Wavefront estimation in astronomical imaging

by

(Ingenieur) Roy Irwan

Department of Electrical and Electronic Engineering

A thesis presented for the degree of
Doctor of Philosophy

University of Canterbury, March 1999

Statement of Originality

This thesis contains no material which has been accepted for the award of any other degree or diploma in any tertiary institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except when due reference is made in the text.

Roy Irwan

Dedicated to my dear parents.

Trustful words are not beautiful.
Beautiful words are not trustful.
Good men do not argue.
Those who argue are not good.
Those who know are not learned.
The learned do not know.

-Lao Tsu (6^{th} century B.C.)

Abstract

The challenge in building astronomical telescopes is to obtain the clearest possible image of a distant star, which should appear as a single point. Extended objects, such as galaxies and planets can be regarded as collections of points.

However, turbulence in the atmosphere degrades any optical signal that passes through it. The optical effects of the atmospheric turbulence arise from random inhomogeneities in the temperature distribution of the atmosphere. As a consequence of these temperature inhomogeneities, the index of refraction distribution of the atmosphere is random. Plane waves striking the atmosphere from space objects acquire an aberration as they propagate through the atmosphere. The plane wave's surface of constant phase is no longer planar when intercepted by an astronomical telescope. The practical consequence of atmospheric turbulence is that resolution is generally limited by turbulence rather than by optical design and quality of a telescope.

There are a number of approaches to solving this problem, ranging from an orbiting telescope (the Hubble Space Telescope), adaptive optics, and post detection processing. The latter approaches have applications to less expensive ground based telescopes and have been the subject of many years of research.

Adaptive optics is a general term for optical components whose characteristics can be modified in real time so as to alter the phase of an incident optical wavefront. An adaptive optics system can be used to correct for atmospheric induced distortions. Before any corrections can be applied, however, some measurement must be made of the phase distortions. It is the aim of this study to estimate the degradation of the wavefronts phase. Two approaches to do so are presented.

Firstly, through wavefront sensors, which many adaptive optics systems have been devised from. Among them the Shack-Hartmann sensor is the most commonly used. The sensor

requires a subdivision of the receiving pupil by means of sub-apertures, wherein the lowest-order deformation of the wavefront phase is estimated. This linearizes the problem of phase retrieval to solving a linear system of equations. A new analysis is presented which differs from previously published work in the estimation of the noise inherent in the centroid calculation used in this sensor. This analysis is supported by computer simulations.

Secondly, the nonlinear approach of phase retrieval is discussed. The problem becomes how to relate the phase and magnitude of the Fourier transform. It is thus necessary to estimate the phase distortion in the instrument solely from measurements made at the image plane of the telescope. The process of phase retrieval is then divided into two distinct steps. The expression for the covariance of the phase distortion using a Kolmogorov model for the turbulence is derived first. This covariance is then employed as part of a formal Bayesian estimate of the phase distortion. It is also shown that phase retrieval can be employed as a robust technique for estimating the wavefront distortion using a lenslet array. The results obtained compare favorably with the alternative approach of phase diversity. Furthermore, the introduction of prior information, in the form of statistical information of the distortion, is shown to considerably enhance the success of the phase retrieval especially for very low light levels.

A comparative evaluation shows the superiority of phase retrieval to Shack-Hartmann sensing, only if the local maxima are overcome. The principal drawback of phase retrieval is the relatively long computing time required to find the solution when general-purposed computer is used.

Acknowledgements

I would like to thank Dr. R. G. Lane under whose supervision of this work took place.

The financial support provided by Dr. Lane and Dr. Bones during my last year is greatly acknowledged.

Finally and most importantly, I wish to express my very special gratitude to my parents and my sisters for all the support they provided during the past years. Without them, this work would have been impossible.

Contents

Preface

In this preface the historical background to my work is outlined and the organization of this thesis is set in context.

Having completed an "ingenieur" degree that is almost equivalent to a Masters degree in electrical and electronic engineering, at Delft University of Technology, the Netherlands, I arrived in March 1996 at the University of Canterbury in New Zealand to undertake postgraduate study under supervision of Dr. R. G. Lane.

I began as a Masters student in which during my first year I completed four useful Masters papers: *Medical Imaging* by Professor T. M. Peters with Dr. Bones' assistance; *Nonlinear Signal Processing* by Dr. Li; and *Signal Processing* II by Dr. Lane and Dr. Li which counted for two Masters papers. I also attended another Masters course: *Adaptive Optics*, by Professor C. J. Dainty, which has supported this work incredibly. Chapter 3 of this thesis is based upon that course. I then transferred to a Ph.D degree in September 1996. Further, I started work on my project of field Astronomical Imaging.

Not surprisingly, Astronomical Imaging is still a major application of digital image restoration today. Ground-based imaging systems were subject to blurring due to the rapidly changing index of refraction of the atmosphere, and I began to investigate the effects of this random inhomogeneities in the temperature distribution of the atmosphere. Plane waves entering the atmosphere from space acquire an aberration as they propagate through the atmosphere. The plane wave's surface of constant phase is no longer planar when received by an astronomical telescope on earth. The practical consequence of atmospheric turbulence is that resolution is generally limited by turbulence. The Hubble Space Telescope is an example of one solution to this problem, but the costs of building and maintaining orbiting telescopes are prohibitive. Adaptive optics and post detection processing such as phase

retrieval and blind deconvolution are approaches that have application to less expensive ground based telescopes and have been the subject of many years of research.

The objective of this thesis is to describe turbulence effects on imaging systems, and to present two essential concepts of wavefront estimation for overcoming the effects of turbulence on imaging systems. Most of the wavefront estimation techniques can be classified into two categories. In the first category, a wavefront is estimated using a wavefront sensor that is a critical component in an adaptive optics. My investigations centered on the most commonly used wavefront sensor: the Shack-Hartmann sensor. As research progressed into the accuracy of the sensor, the assumption in the previously published work that image spots formed by this sensor had a Gaussian profile was shown to be inaccurate. We then examined the correct formulation of the measurement noise in the sensor.

In the second, the wavefront is estimated by use of a nonlinear approach of phase retrieval. The estimation of wavefront through phase retrieval can be reformulated as an inverse problem which is generally ill-posed. Following a suggestion from Dr. Lane, the incorporation of a priori information regarding the atmospheric turbulence into the estimation was developed so that the problem became well-posed. The method has a better potential performance than that of the Shack-Hartmann sensor, but the some of the issues of robustness are still unsolved. This work has been reported in the recent publication [69].

The result of my Ph.D (and the like) studies have led to my accepting a position of Research Scientist at *Philips Research Laboratories* in Eindhoven, the Netherlands.

0.0.1 Thesis organization

This thesis is organized into seven chapters. It is written in such a way it groups together common topics into single chapters accompanied with experimental work. However, only brief conclusions are drawn in these chapters as a single comprehensive conclusion is reserved to the last chapter.

The first chapter gives a brief introduction to the aspect of imaging in general context, and astronomical imaging including the problems and solutions in particular.

In chapter 2, the notation and mathematical relationships which are required in this thesis

are introduced. The Fourier transform and the fundamentals of statistics and random processes are presented. These two sets of background are the basis of the work in this thesis.

Chapter 3 is a review of relevant aspects of the optical effect of atmospheric turbulence. This chapter is based on a Masters course lecture given by Professor Dainty from Imperial College, London.

Chapter 4 extends the detail of the previous chapter, concentrating on the wavefront sensing. The background theory of three existing sensors are given and a new analysis of a particular wavefront sensor, namely the Shack-Hartmann sensor is presented.

Chapter 5 presents a new approach of phase retrieval in astronomy. It uses a prior knowledge which is derived from the statistics of the atmospheric turbulence based on the Kolmogorov model. A combination between phase retrieval and the Shack-Hartmann approach, resulting in multi-aperture phase retrieval is also described.

Chapter 6 compares the performance of the Shack-Hartmann sensing with its new analysis with phase retrieval discussed in chapter 5.

Finally, an overall conclusion, a summary of a number of points of interests, and the major results of this thesis are discussed in chapter 7. Also included in this chapter a brief discussion of possible future extensions of this work.

0.0.2 Supporting publications

This research has resulted in two journal papers and a number of conference publications. These are listed below:

- **R. Irwan** and R. G. Lane, "Phase retrieval with prior information", Journal of the Optical Society of America A **15**, 2302-2311, (1998).

- **R. Irwan** and R. G. Lane, "An analysis of optimal centroid estimation applied to Shack-Hartmann sensing", *submitted to* Applied Optics, (1999).

- R. G. Lane, R. A. Johnston, **R. Irwan**, T. J. Connolly, "Regularized blind deconvolution", Technical Digest Series **11**, Optical Society of America, Hawaii, 5-7, (1998).

- P. J. Bones, R. G. Lane, **R. Irwan**, "Effects of truncation on deconvolution", Proceedings of SPIE, San Diego, **3171**, (1997).
- R. G. Lane, **R. Irwan**, "Phase retrieval as a means of wavefront sensing", Proceedings of International Conference on Image Processing, IEEE Signal Processing Society, Santa Barbara, **2**, 242-245, (1997).
- P. J. Bones, **R. Irwan** and R. G. Lane, "Scanning SVD filters for deconvolution", Digital Image Computing: Techniques and Applications, Albany, 171-176, (1997).
- **R. Irwan** and R. G. Lane, "Phase retrieval using prior information", Proceedings of Image Vision and Computing New Zealand, Lower Hutt, 137-140, (1997).

Glossary

0.0.3 Symbols and abbreviations

$\leftrightarrow$	Fourier transform pair
$P_r\{x\}$	Probability of x
$\delta(i,j)$	The Kronecker delta function, equal 1 if $i=j$ and 0 otherwise.
$\approx$	Approximately equal to
$\propto$	Proportional to
∞	Infinity
$\sum$	Summation
$\prod$	Product
$\langle\rangle$	ensemble average
A^{-1}	Inverse of the matrix A
A^T	Matrix transpose of A
σ^2	variance
$\odot$	Convolution operator
$*$	Correlation operator
π	Pi (3.1415...)
i	$\sqrt{-1}$
r_0	Fried's parameter
x^*	Complex conjugate of x
$\|x\|$	Absolute value of x

0.0.4 Terms

- *Angular resolution:*
  A telescope's ability to distinguish, or resolve an adjoining pair of objects (such as double stars) into two separate images.
- *Coherent imaging:*
  An imaging system where the object is illuminated by laser light.
- *Incoherent imaging:*
  An imaging system where the object is either illuminated by thermal light or self-luminant.
- *CCD:*
  Charge Coupled Device.
- *CDF:*
  Cumulative Density Function.
- *CG:*
  Conjugate Gradient.
- *D:*
  The aperture dimension of a telescope.
- $\mathcal{D}$
  Structure function.
- *DFT:*
  Discrete Fourier Transform.
- $\mathcal{F}$
  Fourier Transform.
- *FFT:*
  Fast Fourier Transform.
- *IFFT:*
  Inverse Fast Fourier Transform.
- *Image plane:*
  A name for the location of a two-dimensional image.

- *PDF:*
  Probability Density Function.
- *Pixel:*
  Picture element - the smallest element of a digital image. The basic unit of which a digital image is composed.
- *MAP:*
  Maximum A Posteriori.
- *ML:*
  Maximum Likelihood.
- *MSE:*
  Mean Square Error.
- *SNR:*
  Signal-to-Noise Ratio.
- *OTF:*
  Optical Transfer Function, is the Fourier transform of PSF.
- *PSF:*
  Point Spread Function, or impulse response of an imaging system.
- *Seeing:*
  The angular extent to which a star's image is spread by turbulence in the earth's atmosphere; good seeing refers to atmospheric conditions in which most of the light is concentrated in a diameter smaller than 1 arc second.
- *Wavefront:*
  A signal found by tracing out an equal optical path from a source to the region of interest, for example, the entrance pupil of an optical system.

Chapter 1

INTRODUCTION

1.1 Aspects of Imaging

Many activities in physics and engineering involve the determination of one physical quantity from measurements of another [10]. To extract information regarding these quantities, here referred to as *objects*, scientific *measurements* are made. The name object is given to the physical quantity because the quantity is often material, such as a star in astronomy or a microbe in microscopy. Measurements are often made directly from the object of interest. In practical situations frequently encountered, however, the object itself is not measurable, and information about it must be inferred from some observable quantity which is related to it. To relate indirect measurements to the properties of the object, the relationship between the two must be first be understood and mathematically modeled. The measured data may then be transformed or inverted to produce data representing an estimate of the object. Such an object estimate is here referred to as an *image*. Resolution, noise level, and measurement artifacts are examples of the factors that may influence one's judgment as to what to do with an image.

In the diverse areas of science and engineering, such as radio engineering, crystallography, microscopy and astronomy, the relationship between the object and the measured data is modeled by the Fourier transform (described in section 2.2). Thus, the measurement is the Fourier transform of the object, and the image is formed by inverse Fourier transformation of the measured data. This relationship is depicted in Fig. 1.1. It should be noted that the last step can be performed using a lens system which collects a portion of this reflected

light and directs it onto the imaging plane where an image is formed.

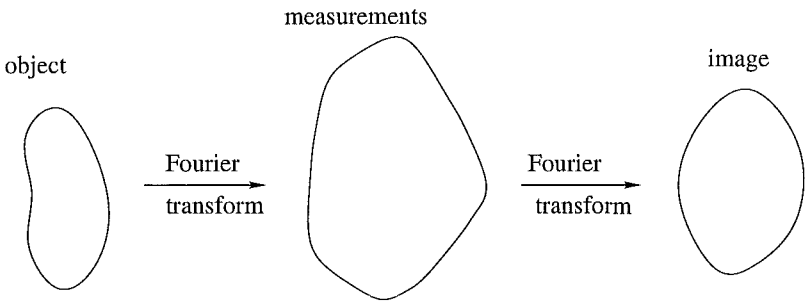


Figure 1.1: *The relationship between objects, measurements, and images.*

It is clear that the measurements affect the image obtained. It should also be noted that only a finite number of measurements can be made in practice due to factors such as limited measurement time or storage capacity. Furthermore, measurements can only be made within a finite region of space. Noise is also inevitably included in measurements, meaning spurious information is recorded which would not present under ideal conditions. The sources of noise are widespread. For example, the astronomical imaging degradation problem is often characterized by photon noise, which is image-dependent and has its roots in the photon-counting statistics involved with low light levels. Another type of noise found in other digital image applications is Gaussian noise, which often arises from the electronic components in the imaging system and broadcast transmission effects [11].

1.1.1 Image model

In this thesis an image, and its Fourier transform are said to exist in two-dimensional coordinate spaces known as image space and Fourier space, respectively. Image space and Fourier space are spanned by the Cartesian position coordinates (x, y) and frequency coordinates (u, v), respectively.

In general, the degradation introduced by the imaging system may be very complicated [11]. However, in many cases of practical importance, such as camera motion, atmospheric turbulence and blur due to the optical transfer function (OTF) of lenses, the imaging system can be modeled as a linear system [12]. Then the degraded image can be modeled by the

following two-dimensional superposition summation [54]:

$$d(x,y) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} f(k,l)h(x,y;k,l) + n(x,y) \tag{1.1}$$

where $f(k,l)$ is the original image, $h(x,y;k,l)$ is the impulse response or point spread function (PSF) of the linear system, and $n(x,y)$ is an additive noise term.

In Eq. (1.1), $h(x,y;k,l)$ can be, in general, spatially varying. The difficulty in solving the recovery problem with a spatially varying blur commonly motivates the use of a stationary model for the blur. This leads to a spatially invariant PSF, i.e. it is independent of position, and therefore Eq. (1.1) reduces to a two-dimensional convolution (see section 2.2.4)

$$\begin{aligned} d(x,y) &= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} f(k,l)h(x-k,y-l) + n(x,y) \\ &= f(x,y) \odot h(x,y) + n(x,y) \\ &= g(x,y) + n(x,y) \end{aligned} \tag{1.2}$$

where the symbol $\odot$ is used to denote the convolution operator. Models that utilize space variant PSFs are also common, but lead to more complicated solutions. Figure 1.2 provides an overview of the linear degradation model for images.

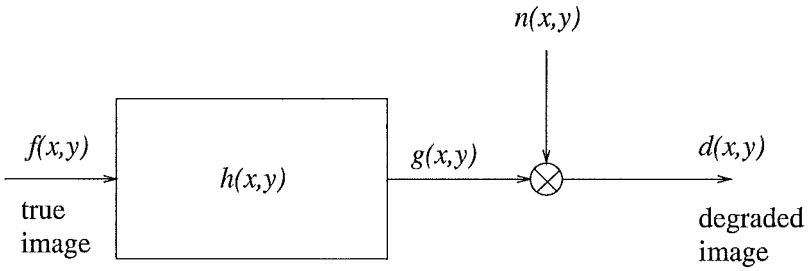


Figure 1.2: *Linear degradation model.*

1.1.2 The image recovery problem

In order to obtain a desired image from observable data it is necessary to consider an inversion process which is often referred to as image recovery. The goal of image recovery is to reconstruct the original object from the degraded observation. This recovery process is critical to many image processing applications such as astronomical speckle imaging [64], remote sensing [25], and medical imaging [26], among others.

The relationship between the original image, a degraded image, and an image estimate is depicted in Fig. 1.3. An ideal image, that is the original image or scene $f(x, y)$, is obtained

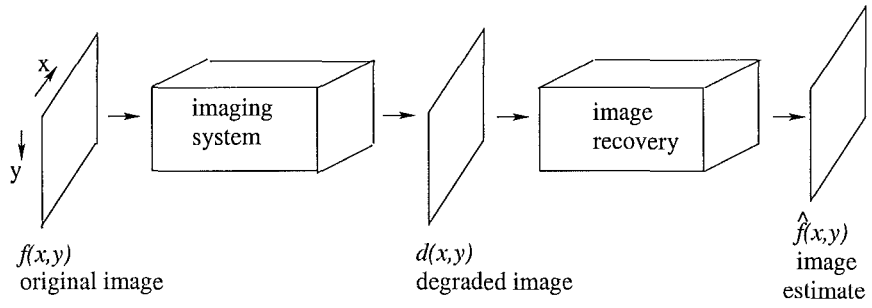


Figure 1.3: *Image recovery setting.*

if the imaging system is perfect. However, since this is not the case, a blurred image $d(x, y)$ is obtained. The goal is to produce from the degraded image $d(x, y)$ an estimate $\hat{f}(x, y)$ which is as close to the original image $f(x, y)$ as possible.

In classical linear image recovery, the blurring function is given, and the degradation process is inverted using one of the many known restoration algorithms. The various approaches that have appeared in the literature depend upon the particular degradation and image model [23, 24]. Unfortunately, in many practical situations, the blur is often unknown, and little information is available about the true image. Blind image recovery is the process of estimating both the true image and the blur from the observed data characteristics, using partial information about the imaging system.

Other practical limitations may cause the measured data in many situations to be incomplete. The measurements are then insufficient to enable the object to be accurately estimated by a simple recovery technique. In such situations, *a priori* information must be used to obtain a useful solution.

1.2 Astronomical Imaging

Not surprisingly, astronomical imaging is still one of the primary applications of digital image recovery today, and most of the work in this thesis deals with this application. The objects of interest in this field can be divided into two classes: passive radiators such as the moon and planets, and active radiators such as stars. In either case the sources of wave

motion are sufficiently distant that a Fourier transform exists between the object and the radiation pattern incident on the earth. This is complicated, however, by the existence of a turbulent atmosphere between earth-bound observers and the objects they wish to observe. The wave motion is distorted by variations in the atmospheric refractive index caused by temperature fluctuations [55]. This distortion affects the ground-based imaging system, and therefore the quality of the observed images. The more details of this type of degradation process and a brief introduction to a real time solution are discussed in this section.

1.2.1 The astronomical setting

Consider the simplified instrument shown in Fig. 1.4, which is forming an image of a distant stellar object. An object or in this case a distant star, which is assumed to be a point source, produces spherical wavefronts (see section 3.1). Because the distance is very large, for all practical purposes the wavefronts entering the earth's atmosphere are planar. At this stage, a good telescope above the atmosphere such as the Hubble Space Telescope is able to form almost perfect or diffraction-limited images. However, after propagating through the random refractive index of the atmosphere, the wavefront entering the telescope pupil is random, and its statistics determine the image quality.

The goal of much current research in this field is to remove the effect of atmospheric turbulence, and although significant progress has been made, there remains much to be done. The challenge in building astronomical telescopes is to obtain the clearest possible image in which case a single star should appear as a single point. Extended objects such as galaxies and planets can be regarded as collections of point sources.

The behaviour of the atmospherically distorted wavefront at the telescope pupil is normally modeled in terms of Kolmogorov theory [4], which predicts the statistical properties of the refractive index fluctuations and leads to equations which describe the propagation of light from the top of the atmosphere to the telescope aperture. The theory predicts that the statistics of the phase fluctuations are fractal [36] and that their statistics are parameterized by a single time-dependent parameter, known as r_0 (discussed in chapter 3). This parameter can be thought of as the maximum diameter of telescope which can support diffraction-limited imaging under the prevailing atmospheric conditions, or "seeing". This parameter is often referred to as the Fried parameter, and it is a measure of how strong the turbulence is. At a good observing site r_0 (in the visible region of the spectrum) can reach

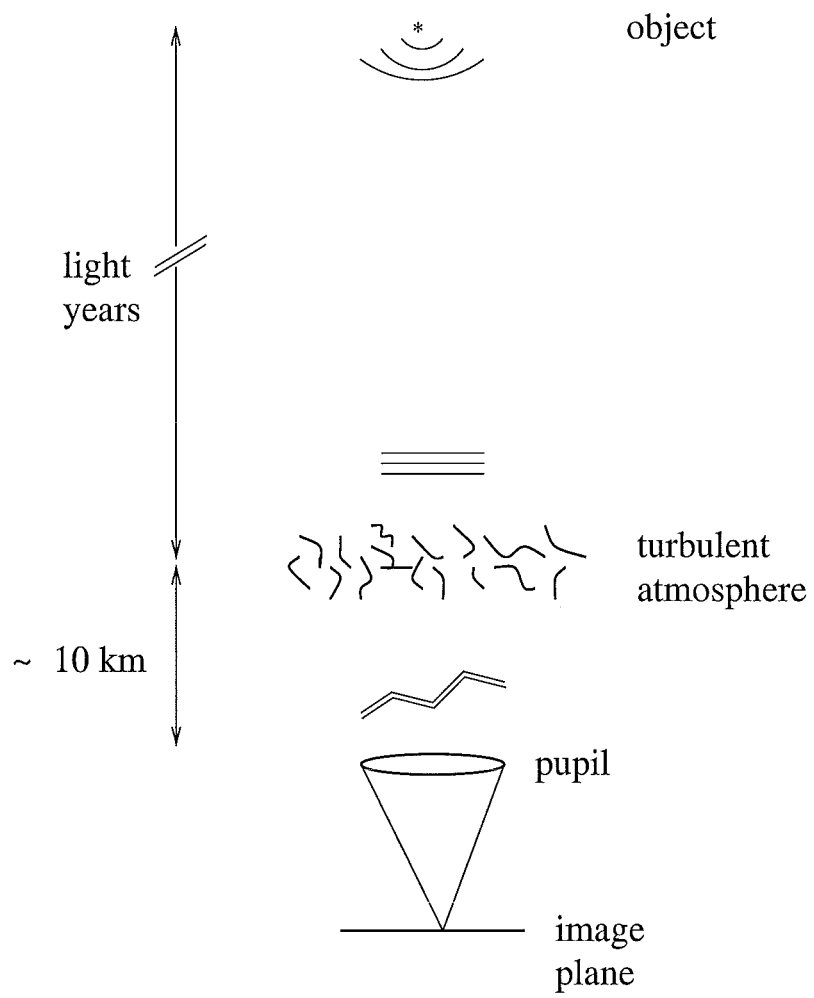


Figure 1.4: *Diagrammatic representation of the astronomical setting.*

20 cm under conditions of good "seeing", whilst under strongly turbulent conditions it may be much less than this.

A frequently raised question is why large ground-based telescopes are being built when atmospheric limitations can be avoided completely by going into space. The answer is that space telescopes are far more expensive to build and maintain than those on the ground. Furthermore, astronomers wish to build telescopes with the largest apertures possible for two reasons: their superior light-collecting capabilities, and their improved potential resolution or diffraction limit. However, atmospheric turbulence imposes a limit on the latter and reduces the performance of a large telescope, in terms of resolution to that of the parameter r_0 at visible wavelengths of light, or worse.

To be more specific, the resolution of large ground-based telescopes is limited by random wavefront aberrations (discussed in the next chapter) caused by atmospheric turbulence. Real time wavefront reconstruction systems, commonly called adaptive optics systems, have been shown to improve the image resolution of these telescopes. A brief introduction to an adaptive optics system for astronomical imaging is now presented.

1.2.2 Adaptive optics

An adaptive optics telescope first suggested by Babcock [2] in 1953 seeks to reverse the effect of image distortion by restoring the spherical shape of the wavefront. The intention is to remove the distortions in the wavefront before the light is detected, forming an improved image in real-time. An adaptive optics system consists of four major constituent parts other than the telescope in which it is incorporated. First, a bright source, such as a bright unresolvable star, must be used to provide a flat reference wavefront which is propagated along the same path and encounters the same turbulence as the object under study. Second, a wavefront sensor is used to measure the distortions introduced by the turbulence into this reference wave. Third, an optical element, which is known as the corrector, is used to remove some of these distortions from the reference wave and also, therefore, from the wave arriving from the science object. This is often a deformable mirror, whose shape can be controlled and rapidly altered with the application of the appropriate signals. The final important constituent of an adaptive optics system is the control algorithm, which must be capable of converting wavefront data taken by the sensor into corrector control signals in a period of time short enough in order that the turbulence has not decorrelated significantly between sensing and correction. Figure 1.5 represents a schematic adaptive optics system.

Aberrated light entering the telescope is first received by the primary mirror, and it is further concentrated into a narrow beam through the secondary mirror. The beam is then reflected from the deformable mirror which is made in segments and fitted together like bathroom tiles. Some of this beam is focused to form an image, and some is used to provide signals to the wavefront sensor. An actuator command computer processes the wavefront sensor's measurements and calculates the set of electrical signals to apply to the deformable mirror to achieve the desired deformable mirror figure. The entire process, from wavefront sensor measurement to the deformable mirror update, must be performed faster than the rate of change of the turbulence.

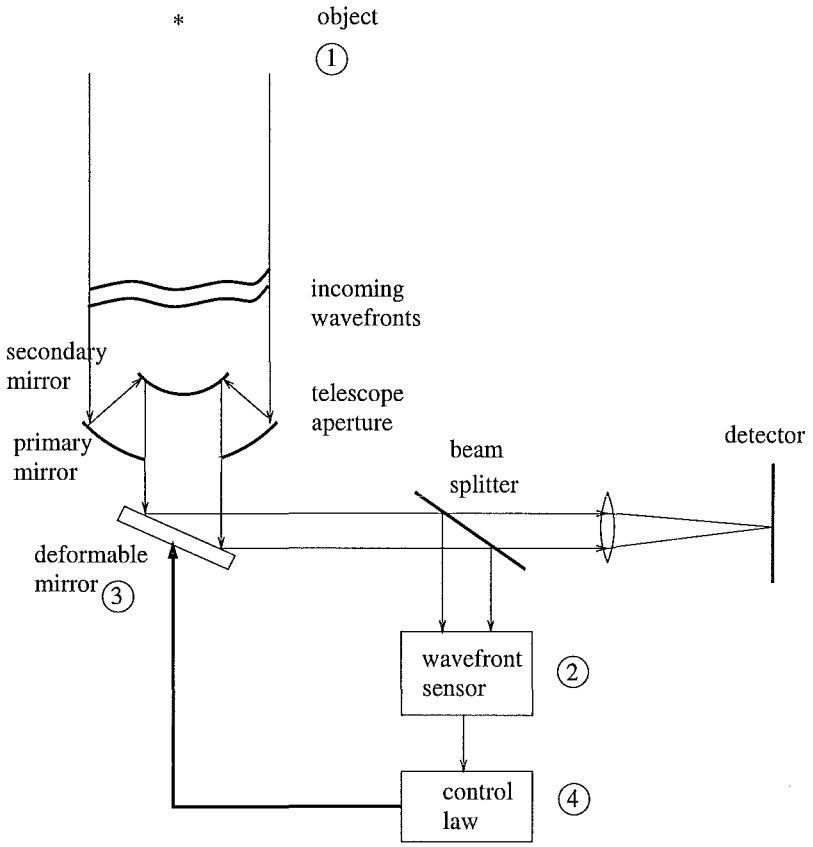


Figure 1.5: *Simplified optical configuration of an adaptive optics system.*

In an ideal adaptive optics system, the surface of the deformable mirror is configured from instant to instant to present an approximation of the conjugate of the turbulence-induced phase error so that the wave reflected from the deformable mirror closely approximates a plane wave. However, it is impossible for an adaptive optics system to fully restore the imaging performance of a ground-based telescope due to finite measurements in the wavefront sensor, finite spatial sampling of the phase aberrations by the wavefront sensor, finite degrees of freedom available in the deformable mirror, and time delays between sensing and correcting for the turbulence-induced aberrations.

1.3 The Wavefront Estimation Problem

This section introduces the problem of wavefront estimation as an essential step in the correction process discussed in section 1.2.2. This problem is the primary topic addressed within the thesis. The section begins with a very brief introduction to wavefront sensing,

which is the current technology used in adaptive optics systems. The phase retrieval problem is formally defined in subsequent section.

1.3.1 The wavefront sensing problem

In this section the problem of estimating wavefront phase from a wavefront sensor is introduced. The reconstruction process is necessary to obtain phase estimates from certain types of sensors, such as a shearing interferometer, a curvature sensor, or a Shack-Hartmann sensor. The shearing interferometer and the Shack-Hartmann sensor measure the wavefront phase slope, permitting the reconstruction of a wavefront phase surface. The curvature sensor is another type of wavefront sensor that, unlike the shearing interferometer and the Shack-Hartmann sensor, senses the curvature of the wavefront phase.

The scope of this thesis is limited to an analysis of the commonly used wavefront sensor, namely the Shack-Hartmann sensor. This sensor requires a subdivision of the receiving pupil by means of lenslet array, wherein the lowest order deformation of the wavefront phase is estimated. Previous work to analyze the performance of this sensor has resulted in disagreements between theory and actual performance [42,43]. One of the reasons includes the assumption of a Gaussian distribution for the illumination at the image plane. Another reason is that the mean number of photons has been used in the sensor measurements instead of the actual number of photons. All these factors can significantly affect the performance of the sensor, as is shown in chapter 4.

Determining how the Shack-Hartmann sensor performance is affected by each of these factors is critical when one is designing an adaptive optics system. Furthermore, the question of how well an adaptive optics system performs under less-than-ideal operating conditions has been the subject of much ongoing research, and a point in contention in this thesis. These less-than-ideal conditions result from the inability to build perfect wavefront sensors and wavefront correction devices. As detailed in chapter 4, the accuracy of a Shack-Hartmann sensor is limited by photon noise and by the finite number of sampling areas over the wavefront surface. Wavefront correction devices are also less than ideal. The ability of a correction device to cancel wavefront aberrations is limited by the finite number of degrees of freedom in the device's response. This limited response prevents it from correcting higher order wavefront aberrations.

For these reasons, this thesis also considers an alternative technique to estimate wavefronts based on phase retrieval, which is now introduced.

1.3.2 The phase retrieval problem

In order to define the phase retrieval problem, it is necessary at this point to introduce the invertible relationship between an image and its spectrum, which is represented by a Fourier pair:

$$f(x,y) \leftrightarrow F(u,v). \tag{1.3}$$

The Fourier transform is central to the techniques described in this thesis and is more fully described in section 2.2. The spectrum is in general complex-valued and thus may be separated into its magnitude $|F(u,v)|$ and phase $\mathrm{Ph}\{F(u,v)\}$ so that

$$F(u,v) = |F(u,v)|\exp[i\mathrm{Ph}\{F(u,v)\}], \tag{1.4}$$

where i is the imaginary operator equal to $\sqrt{-1}$. The Fourier *intensity* is defined as

$$|F(u,v)|^2 = F(u,v)F^*(u,v) \tag{1.5}$$

where the superscript $*$ denotes complex conjugation. The Fourier intensity forms a Fourier pair with the autocorrelation $ff(x,y)$ of $f(x,y)$:

$$ff(x,y) \leftrightarrow |F(u,v)|^2. \tag{1.6}$$

If one wants to reconstruct an image $f(x,y)$, both magnitude and phase of $F(u,v)$ are usually required. As can be seen from Eq. (1.6), the information contained in the square magnitude of the Fourier transform of $f(x,y)$ is the same as that contained in the autocorrelation function. Since the magnitude of a function must be real, it contains no phase information in the Fourier phase. This implies that the phase information about $F(u,v)$ is also completely lost in its autocorrelation function. The phase retrieval problem is the task of retrieving an estimate of the true image when the magnitude (or equivalently the intensity) of the spectrum is known accurately and the Fourier phase is either partially or totally unknown [72]. The problem is so named because the unknown information regarding the Fourier phase must be retrieved.

It is useful to further divide the phase retrieval problem into two types. In many situations an estimate of the Fourier phase and/or the true image is measurable. The task of recovering

the true image from the accurate Fourier magnitude and this additional information is referred to as a *partial phase retrieval problem* [72]. The more difficult and less frequently occurring problem, where only the Fourier magnitude or intensity is available, is called the *pure phase retrieval problem* [72]. Although more difficult, the pure phase retrieval problem provides useful insight into the interrelationship between Fourier magnitude and phase. This thesis only considers solving the pure phase retrieval problem because any techniques that effectively solve the pure phase retrieval problem should provide the basis for solutions to the partial phase retrieval problem. They may also be usefully adapted to incorporate the additional information available in a manner covered in chapter 5.

The pure phase retrieval problem is equivalent to recovering the true image from its autocorrelation $ff(x,y)$, since the Fourier intensity is the Fourier transform of the autocorrelation. However, the autocorrelation is, in practice, larger than $f(x,y)$, therefore in order to compute the autocorrelation correctly from the Fourier intensity, measurements (i.e. samples) of the latter must be finely enough spaced in Fourier space. For this to be the case, the Fourier magnitude must be *oversampled* (see section 2.2) by a factor of at least 2, and the phase retrieval problem is then referred to as a *Fourier phase retrieval* [72]. The new phase retrieval technique introduced in chapter 5 is only applicable to Fourier phase retrieval problems. For convenience, in this thesis the Fourier phase retrieval problem is simply called phase retrieval.

A major problem with the phase retrieval is the issue of uniqueness. If no *a priori* knowledge about the true image is available, then phase retrieval has no unique solution, since there is no way of choosing one Fourier phase over another. In practice, however, it is often known that the amplitude of the true image is non-zero over only a finite region of image space, known as the image *support* (section 2.2). In any pure phase retrieval problem, trivial ambiguities always exist: the location and orientation in image space of the true image cannot be recovered since shifted and/or conjugated and 180 degrees rotated forms of the true image all have the same Fourier magnitude. However, all these trivial ambiguities have basically the same appearance or "form", which is referred to as the *image form* [5].

The importance of phase information can be demonstrated by the following example. Consider an image obtained by the Hubble Space Telescope [73] shown in Fig. 1.6. The Fourier transform of the image is shown in Fig. 1.7. If the spectral magnitude of the Fourier transform of the image is replaced by a constant and the phase is kept unchanged, the resultant

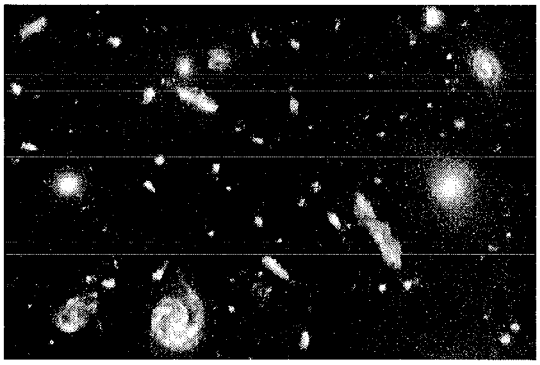
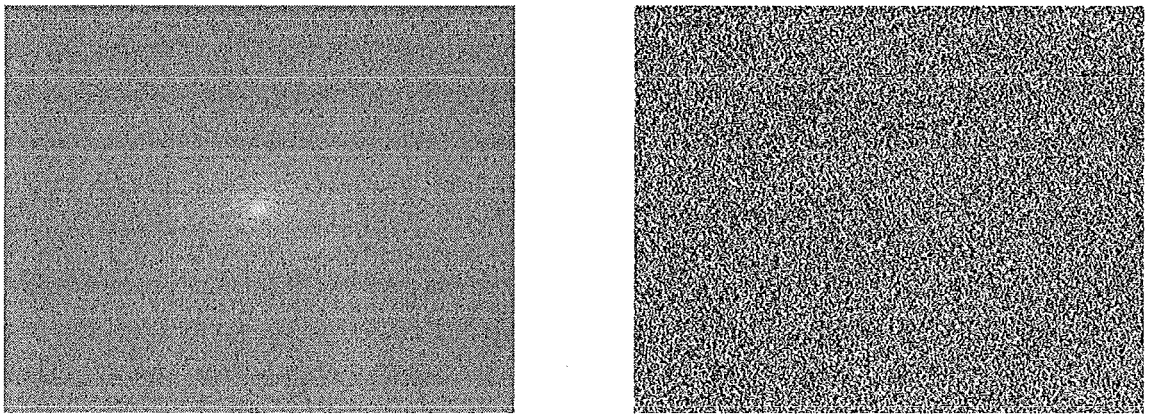


Figure 1.6: *The image used to demonstrate the importance of the phase information in image recovery.*

image is shown in Fig. 1.8. If now the phase of the Fourier transform of the image is replaced with some random numbers but the spectral magnitude is kept unchanged, the resultant image is shown in Fig. 1.9.



(a) (b)

Figure 1.7: *(a) The magnitude on a log scale and (b) the phase of the Fourier transform of Fig. 1.6.*

Comparing the images shown in Fig. 1.6, Fig. 1.8, and Fig. 1.9, the image that retains the Fourier phase can still be recognized. However, the shape of the image that retains the Fourier magnitude changes and is significantly different from the original one. Phase information is thus more important than the magnitude information in recognizing and also reconstructing an image [11].

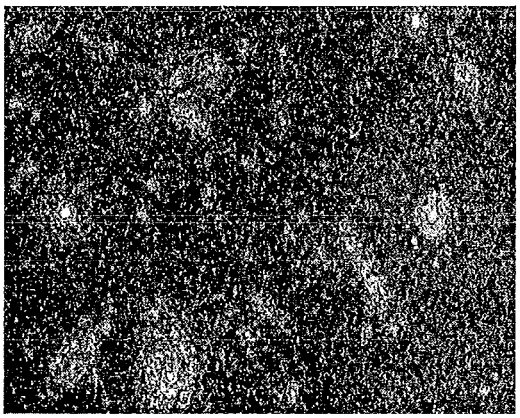


Figure 1.8: *The reconstructed image formed by replacing all the spectral magnitude to 1 but keeping the phase information unchanged.*

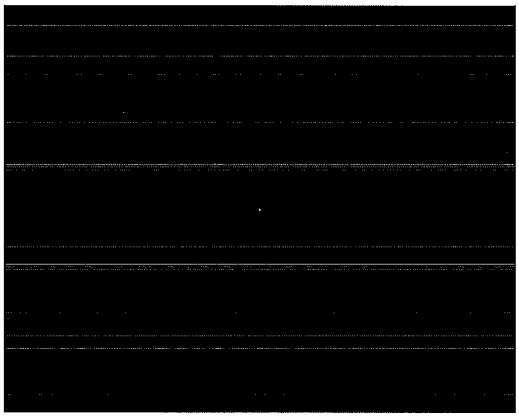


Figure 1.9: *The reconstructed image formed by replacing all the phase to a set of random numbers but keeping the magnitude information unchanged.*

Chapter 2

PRELIMINARIES

The purpose of this chapter is to introduce the notation, terminology, and mathematical tools necessary to enable exposition in the remainder of this thesis.

The first section of this chapter introduces the mathematical notation used to represent quantities such as vectors, matrices, functions, and complex numbers. A brief review of the Fourier transform is addressed in section 2.2. The properties of the Fourier transform together with a practical knowledge of their physical interpretation is fundamental to approach most image processing problems. When performing computations on a digital computer, it is often necessary to sample both the image and its Fourier transform. The discrete Fourier transform, which relates samples in image and Fourier space is described and its relationship to the Fourier transform is explained. An introduction to statistics and random processes is particularly useful in most of the aspects in this thesis and is described in chapter 2.3.

2.1 Notation

Throughout this thesis, vectors are denoted by bold lower case and the n^{th} component of the vector $\mathbf{a}$ is $\mathbf{a}_n$. In a similar manner matrices are represented by a bold upper case and the individual elements of the matrix are signified by lower case subscripts. The transpose and inverse of a matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$ and $\mathbf{A}^{-1}$, respectively.

Functions of variable(s) are indicated by enclosing the variable(s) in parentheses, therefore, $f(x)$ is a function of the variable x. When a function operates on the result of another functional operator this is expressed by placing the first function within a set of curly brackets. Hence, the result of $g(y)$ operating on $f(x)$ is $g\{f(x)\}$. When a function is not known accurately or has been estimated this is indicated by placing a circumflex over the quantity, for instance, $\hat{f}(x)$.

In this thesis, the image space is represented by functions such as $f(\mathbf{x})$. The lower case functional label, f, is used to indicate that the image is in image space and the variable $\mathbf{x}$ is a coordinate in this K-dimensional space. When the Fourier transform (see Section 2.2) is used to map images from image space into Fourier space this is indicated by using an upper case functional label and the K-dimensional Fourier space variable $\mathbf{u}$, i.e. $F(\mathbf{u})$ is the Fourier transform or spectrum of $f(\mathbf{x})$. As mentioned in section 1.1, images in this thesis exist primarily in a two-dimensional space, and hence the images in image and Fourier space may also be given in the form $f(x,y)$ and $F(u,v)$, respectively.

Notations related to complex numbers are now defined. The real and imaginary components of a complex number c have been given the notation $\mathrm{Re}(c)$ and $\mathrm{Im}(c)$, respectively. Thus,

$$c = \mathrm{Re}(c) + i\mathrm{Im}(c). \tag{2.1}$$

The complex number c can also be expressed in its polar form

$$c = |c|\exp[i\mathrm{Ph}(c)], \tag{2.2}$$

where $|c|$ and $\mathrm{Ph}(c)$ are the magnitude and phase of c, respectively. The magnitude is given by the expression

$$|c| = \sqrt{\mathrm{Re}(c)^2 + \mathrm{Im}(c)^2}, \tag{2.3}$$

and the phase by

$$\mathrm{Ph}(c) = \tan^{-1}\frac{\mathrm{Im}(c)}{\mathrm{Re}(c)}. \tag{2.4}$$

The complex conjugate of c is denoted by c^*, where

$$c^* = \mathrm{Re}(c) - i\mathrm{Im}(c) = |c|\exp[-i\mathrm{Ph}(c)]. \tag{2.5}$$

2.2 The Fourier Transform

In this section the Fourier transform is formally defined and reasons are given for its applicability throughout a wide range of disciplines. The Fourier transform is a powerful tool

in linear system analysis. In linear systems any signals can be represented as linear combinations of basic signals such as sine waves. The Fourier transform is of importance for the linear system analysis because it provides the basic signals in terms of complex exponential signals in continuous and discrete time, i.e., signals of the form e^{st} in continuous time and z^n in discrete time, where s and z are complex numbers. Furthermore, it is usually used as a mathematical or physical tool to transform a problem into a form that can be easily solved. In this thesis it is sufficient to use this transform and its discrete form, and not the Laplace transform or the Z transform.

2.2.1 Continuous Fourier transform

The Fourier transform of a continuous one-dimensional function $f(t)$ is defined as [77],

$$\mathcal{F}\{f(t)\} = F(\omega) = \int_{-\infty}^{\infty} f(t)\exp[-i2\pi\omega t]\mathrm{d}t, \tag{2.6}$$

where $\mathcal{F}\{\cdot\}$ denotes the Fourier transform operation. The inverse Fourier transform of $F(\omega)$ is defined as,

$$\mathcal{F}^{-1}\{F(\omega)\} = \int_{-\infty}^{\infty} F(\omega)\exp[i2\pi\omega t]\mathrm{d}\omega, \tag{2.7}$$

where $\mathcal{F}^{-1}\{\cdot\}$ denotes the inverse Fourier transform. Furthermore, since

$$f(t) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(t)\exp[-i2\pi\omega t]\mathrm{d}t\right]\exp[i2\pi\omega t]\mathrm{d}\omega, \tag{2.8}$$

this means that the transformation is reciprocal, and

$$\mathcal{F}\{f(t)\} = F(\omega) \Rightarrow \mathcal{F}^{-1}\{F(\omega)\} = f(t). \tag{2.9}$$

For any function $f(t)$, there is a unique one to one relationship to its Fourier transform $F(\omega)$.

2.2.2 Discrete Fourier transform

The Discrete Fourier transform (DFT) can be viewed simply as a way of approximating the continuous Fourier transform. By definition a DFT of a sequence $f(n)$ is given by

$$F(m) = \frac{1}{N}\sum_{n=0}^{N-1} f(n)\exp[-i\frac{2\pi mn}{N}]. \tag{2.10}$$

and the inverse DFT is

$$f(n) = \sum_{m=0}^{N-1} F(m)\exp[i\frac{2\pi mn}{N}], \tag{2.11}$$

where m and n are integer-valued indices. The two equations together form a discrete transform pair, and approximate the continuous Fourier transform pair. The output of the standard DFT algorithm is a set of complex numbers which relate to the magnitude and phase of the sampled sinusoidal components which make up the signal.

Practical application of the DFT involves using the Fast Fourier Transform (FFT), which is a computationally efficient method of computing the DFT. The FFT is most commonly used for sequences when N is an integer power of 2. Although fast algorithms exist for other sequence lengths they are significantly more complicated to implement. For a one dimensional signal, the computational saving of the FFT in this case over the DFT is a factor of $N/\log_2 N$. All examples in this thesis are arranged to meet this size constraint, and the FFT is used for all such computations.

2.2.3 The Fourier transform in two dimensions

In digital image processing the inputs and outputs are commonly two-dimensional or higher, and it is straightforward to extend the Fourier transform to two dimensions. In this case the time variable t is replaced by the two-dimensional spatial variables (x, y) and the frequency f is replaced by the two-dimensional spatial variables (u, v). Thus, the two-dimensional forward and inverse Fourier transform are respectively defined as

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\exp[-i2\pi(ux+vy)]\mathrm{d}x\mathrm{d}y \tag{2.12}$$

and

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)\exp[i2\pi(ux+vy)]\mathrm{d}u\mathrm{d}v \tag{2.13}$$

where $f(x, y)$ is a two-dimensional signal and $F(u, v)$ is its spectrum.

As in one dimension, the DFT is quite similar to the continuous Fourier transform. The two-dimensional DFT of an $N \times N$ pixel image is given by

$$F(u,v) = \frac{1}{N^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\exp[-i2\pi(ux+vy)/N] \tag{2.14}$$

and the inverse relationship by

$$f(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} F(u,v)\exp[i2\pi(ux+vy)/N]. \tag{2.15}$$

These formulae are both two-dimensional summations and it is possible to rewrite the transform as

$$F(u,v) = \frac{1}{N}\sum_{x=0}^{N-1}\left[\frac{1}{N}\sum_{y=0}^{N-1} f(x,y)\exp[-i2\pi vy/N]\right]\exp[-i2\pi ux/N] \tag{2.16}$$

thereby separating the transformation into a series of horizontal and vertical operations on rows and columns. The inverse DFT is likewise separable.

2.2.4 Convolution and correlation

Convolution and correlation have been found to be extremely valuable in engineering, where they have been used extensively to facilitate the development of mathematical models of physical systems, particularly linear systems. In this section the mathematical formulae which describe both the convolution and correlation processes are presented and the concepts behind the formulae and their application to system modeling are discussed.

Of the two processes, the process that primarily forms the basis of many mathematical models of a system is convolution. Before introducing the convolution operator it is convenient to introduce the Dirac delta function which is defined by [34, 77],

$$\delta(\mathbf{x}) = \begin{cases} \infty & \text{for } \mathbf{x} = 0 \\ 0 & \text{otherwise,} \end{cases} \tag{2.17}$$

and

$$\int_{-\infty}^{\infty} \delta(\mathbf{x})\mathrm{d}\mathbf{x} = 1. \tag{2.18}$$

A useful property of the delta function is the shifting property

$$\int_{-\infty}^{\infty} f(\mathbf{x})\delta(\mathbf{x}-\mathbf{x}_0)\mathrm{d}\mathbf{x} = f(\mathbf{x}_0) \tag{2.19}$$

which allows $f(\mathbf{x}_0)$ to be considered as the summation of an infinite sequence of contiguous delta functions with each individual delta function being multiplied by the value of $f(\mathbf{x}_0)$ corresponding to the origin of that delta function.

Consider the imaging system depicted in Fig. 2.1 (cf. Fig. 1.2). Assume further that the imaging system is a linear system, represented mathematically by the symbol $\mathcal{L}$, that produces an output image $g(\mathbf{x})$ when an input image $f(\mathbf{x})$ is applied to the system. A linear system is one for which the principle of superposition applies [23], that is:

$$\mathcal{L}\{af_1(\mathbf{x}) + bf_2(\mathbf{x})\} = a\mathcal{L}\{f_1(\mathbf{x})\} + b\mathcal{L}\{f_2(\mathbf{x})\}, \tag{2.20}$$

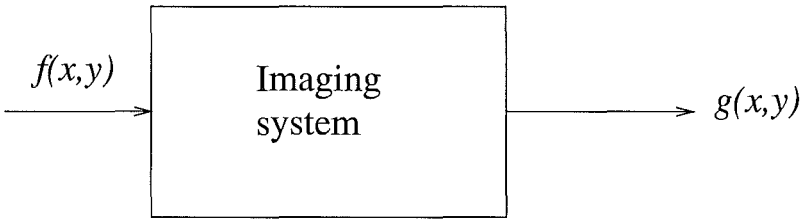


Figure 2.1: *Imaging system.*

where a and b are arbitrary constants. Using the useful shifting property of the delta function in Eq. (2.19), any input can be decomposed into the sum of delta functions. It is thus possible to express the output image as

$$\begin{aligned} g(\mathbf{x}) &= \mathcal{L}\{\int_{-\infty}^{\infty} f(\mathbf{x}_0)\delta(\mathbf{x}-\mathbf{x}_0)\mathrm{d}\mathbf{x}_0\} \\ &= \int_{-\infty}^{\infty} f(\mathbf{x}_0)\mathcal{L}\{\delta(\mathbf{x}-\mathbf{x}_0)\}\mathrm{d}\mathbf{x}_0 \\ &= \int_{-\infty}^{\infty} f(\mathbf{x}_0)h(\mathbf{x},\mathbf{x}_0)\mathrm{d}\mathbf{x}_0 \end{aligned} \tag{2.21}$$

where $h(\mathbf{x},\mathbf{x}_0)$ is the system response to a delta function at position $\mathbf{x}_0$, and is called the PSF or impulse response.

An important class of linear systems is *shift* or *spatially* invariant systems, that is the output is shifted by the same amount as the inputs, but is otherwise unchanged. In this case the PSF may be written as

$$h(\mathbf{x},\mathbf{x}_0) = h(\mathbf{x}-\mathbf{x}_0). \tag{2.22}$$

In practice, systems are not completely invariant but often invariance may be assumed [22–24]. Substitution of Eq. (2.22) into Eq. (2.21) reduces the latter to the more simple process,

$$\begin{aligned} g(\mathbf{x}) &= \int_{-\infty}^{\infty} f(\mathbf{x}_0)h(\mathbf{x}-\mathbf{x}_0)\mathrm{d}\mathbf{x}_0 \\ &= f(\mathbf{x}) \odot h(\mathbf{x}) \end{aligned} \tag{2.23}$$

which is known as convolution. The symbol $\odot$ is used as the convolution operator as introduced in section 1.1.1. In other words, one of the functions is rotated 180 degrees about the origin by reversing the sign of both x and y, displaced, and multiplied with the other function, and the product is then integrated to obtain the value of the convolution integral for that particular displacement. One of the desirable properties of the Fourier

transform is that on transforming both sides of Eq. (2.23), the convolution now becomes multiplication

$$G(\mathbf{u}) = F(\mathbf{u})H(\mathbf{u}) \tag{2.24}$$

where $H(\mathbf{u})$ is known as the *transfer function* of the system.

Correlation is primarily used to determine how two signals are related to each other mathematically. If the correlation function peaks for a particular value of $\mathbf{x}$, this would indicate a very good correlation, which means that the two signals match each other best when there is a displacement of $\mathbf{x}$ between them. Conversely, a very small or zero value of the correlation function indicates little or no correlation. The correlation of two individual signals, known as cross-correlation function, is formed in the same way as convolution but the sign reversal is omitted,

$$f(\mathbf{x}) * h(\mathbf{x}) = \int_{-\infty}^{\infty} f(\mathbf{x}_0)h(\mathbf{x} + \mathbf{x}_0)\mathrm{d}\mathbf{x}_0 \tag{2.25}$$

where $*$ denotes the process of correlation [34].

The auto-correlation function can be considered as a special case of the cross-correlation function with $f(\mathbf{x}) = h(\mathbf{x})$,

$$f(\mathbf{x}) * f(\mathbf{x}) = \int_{-\infty}^{\infty} f(\mathbf{x}_0)f(\mathbf{x} + \mathbf{x}_0)\mathrm{d}\mathbf{x}_0 \tag{2.26}$$

The Fourier transform of the cross and auto-correlation functions often provides useful interpretations of the nature of the signals. It can be shown that the Fourier transform of the auto-correlation function is the square of the magnitude of the Fourier transform of the signal $f(\mathbf{x})$. This function is called the power spectral density function or power spectrum, and it is defined as

$$\begin{aligned} P_{ff} = \mathcal{F}\{f(\mathbf{x}) * f(\mathbf{x})\} &= F(\mathbf{u})F(-\mathbf{u}) \\ &= F(\mathbf{u})F^*(\mathbf{u}) = |F(\mathbf{u})|^2. \end{aligned} \tag{2.27}$$

If $f(\mathbf{x})$ is real, its autocorrelation function is real and even, and therefore, its power spectrum is real and even [21]. Again, any $f(\mathbf{x})$ has a unique power spectrum but the converse is not necessarily the case.

2.2.5 Functions with circular symmetry

A function $f(\mathbf{x})$ is said to be circularly symmetric if it can be written as a function of r alone, where r is the radial distance from the origin [77], that is

$$f(r,\theta) = f_R(r). \tag{2.28}$$

Such functions are particularly of importance in this thesis since most optical systems have precisely this type of symmetry.

If $f(\mathbf{x})$ has circular symmetry, its transform is expressible as $F(\mathbf{u}) = F(\rho)$, where ρ is the radial spatial frequency variable, also has circular symmetry. $F(\rho)$ and $f_R(r)$ are linked by another cyclical transform called the Hankel transform of zero order [77]:

$$F(\rho) = 2\pi \int_0^\infty f_R(r) J_0(2\pi\rho r) r \, \mathrm{d}r \tag{2.29}$$

$$f_R(r) = 2\pi \int_0^\infty F(\rho) J_0(2\pi\rho r) \rho \, \mathrm{d}\rho \tag{2.30}$$

where J_0 is the Bessel function of order zero. Thus, for circularly symmetric functions there is no difference between the transform and inverse-transform operations. An example of zero-order Hankel transform is shown as two-dimensional Fourier transform in Fig. 2.2.

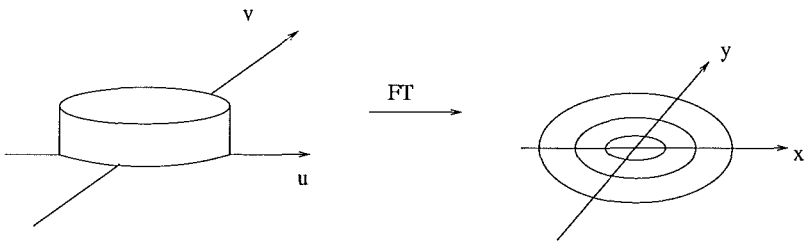


Figure 2.2: *A zero-order Hankel transforms shown as a two-dimensional Fourier transform.*

It is important to note that this transform is a one-dimensional linear integral transform similar to the Fourier transform, except that the kernel is a Bessel function. Hence, two-dimensional functions with circular symmetry may be treated as one-dimensional functions of a single radial variable if the Hankel transform is substituted for the Fourier transform.

2.2.6 Compact images

All real world (i.e. physically realisable) images are of finite extent. The size of the object may be small enough to be within the field of view of the instrument. When the object is

larger, however, the image is truncated, because the field of view is finite in any practical instrument. Therefore most of the information of the object is concentrated into a finite region of space denoted its *support*, S. Furthermore, all real world images contain noise, making it pointless to represent detail below the level of noise or the measurement uncertainty level. Denoting the level of these uncertainties by the small positive real number ε, then the support of an image $f(x, y)$ is defined as the region for which

$$|f(x, y)| > \varepsilon. \tag{2.31}$$

It is convenient to define an *image box* $B(f)$ which has sides parallel to the Cartesian axes and which is just large enough to encompass the support of the image. The term *extent* is used to denote the size of the image box, i.e. the linear dimension of the image $(L_k(f))$. The support, image box, and image extents of a two-dimensional image are illustrated in Fig. 2.3.

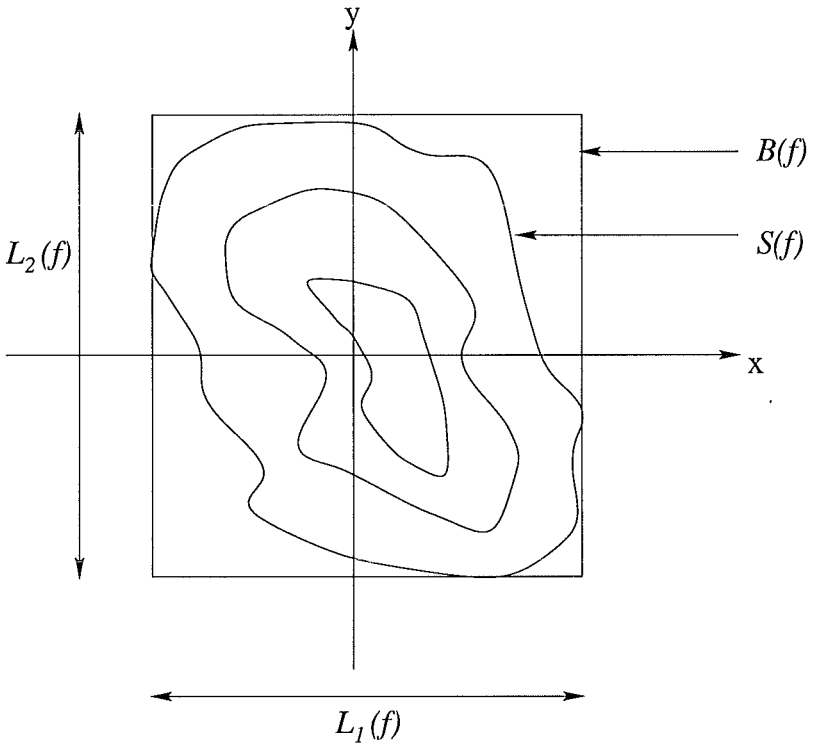


Figure 2.3: *Example of image support, image box, and extents for a two-dimensional image* $f(x, y)$.

If $f(x, y)$ is of finite extent and contains no infinities, which is true for all real world images,

i.e.

$$|f(x,y)| < \infty, \quad \text{for} \quad (x,y) \in B(f) \tag{2.32}$$

then $f(x,y)$ is called *compact* [72]. A compact image is said to be *exactly compact* if $\varepsilon = 0$ in Eq. (2.31), and *approximately compact* otherwise. In order to mathematically model physical situations, it is sometimes useful to assume an image is exactly compact. As a consequence of Parseval's theorem which states that the energy of an image equals the energy of its spectrum, the Fourier transform of a compact image must itself be effectively compact.

2.2.7 Sampling

Since computers can process only digital images, a prerequisite for digital image processing is the conversion of images into digital form. It is thus necessary to represent the image by a finite number of *samples*. In Section 2.2.4 it was demonstrated that sampling a continuous function, by multiplying the function with a series of delta functions, has the effect of convolving the spectrum of the function with the spectrum of a series of delta functions. This series of delta functions (Fig. 2.4) is often called the *comb* function, and is defined as

$$\text{comb}(\mathbf{x}) = \sum_{n=-\infty}^{\infty} \delta(\mathbf{x} - \mathbf{x}_n). \tag{2.33}$$

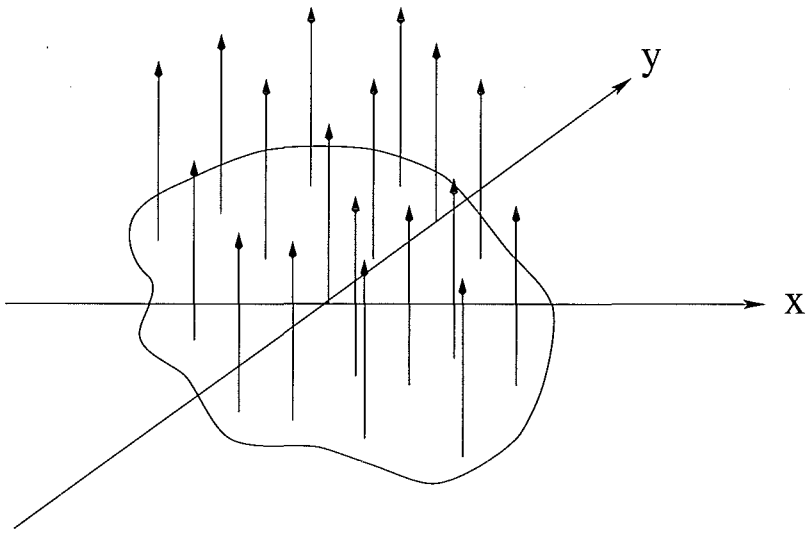


Figure 2.4: *The sampled function.*

According to the Nyquist criterion, if the image or signal, which the continuous function represents, has a bandwidth BW then the sampling period T_s must be less than $\frac{1}{2BW}$, otherwise the individual continuous signal spectra, which form the sampled signal's spectrum, overlap in Fourier space. The overlap in Fourier space superimposes high frequency components on to low frequency components, thereby, distorting the resulting spectrum so that one period of the sampled signal's spectrum is no longer equivalent to the continuous signal's spectrum, an effect that is known as aliasing. Provided that $T_s < \frac{1}{2BW}$, or equivalently if the overlap in Fourier space occurs below the level of the inevitable noise, the effect of aliasing is negligible and the sampled signal is an accurate representation of the continuous signal.

2.2.8 Summary of Fourier transform properties

The summary is given here as they are required throughout this thesis to enable other relationships to be developed or important concepts to be introduced. A number of properties of the Fourier transform are summarized in table 2.1 as they are required throughout this thesis to enable other relationship to be developed. Two properties of the Fourier transform listed in table 2.1, the convolution and correlation theorems, are used extensively throughout this thesis to simulate degraded speckle images which can be modeled as a convolution between point sources and the blurring function caused by the turbulence. Parseval's theorem is often useful in determining some time domain characteristics of a signal directly from the Fourier transform (section 4.2.1). The differentiation property is merely used in the derivation of the angle-of-arrival in this thesis (section 3.2).

Derivations of the properties in table 2.1 can be found in a number of general texts dealing with either Fourier transforms, communications or signal processing [34, 77].

2.3 Statistics and Random Processes

In this section various definitions of probability and their roles in this thesis are presented. Much of the work in this thesis relies heavily on these statistical theorems.

Property	Time or spatial domain	Frequency domain
	Signal	Spectrum
	Impulse response	Transfer function
Terminology	Auto-correlation	Power spectrum
	Cross-correlation	Cross power spectrum
	$f(t)$	$F(\omega)$
Definition	$\int_{-\infty}^{\infty} F(\omega)\exp[i2\pi\omega t]\mathrm{d}\omega$	$\int_{-\infty}^{\infty} f(t)\exp[-i2\pi\omega t]\mathrm{d}t$
Linearity	$af(t) + bh(t)$	$aF(\omega) + bH(\omega)$
Similarity	$f(at)$	$\frac{1}{\lvert a\rvert}F\left(\frac{\omega}{a}\right)$
Shift	$f(t-a)$	$\exp[-i2\pi a\omega]F(\omega)$
Convolution	$f(t) \odot h(t)$	$F(\omega)G(\omega)$
Correlation	$f(t) * h(t)$	$F^*(\omega^*)G(\omega)$
Differentiation	$\frac{d}{dt}f(t)$	$i2\pi\omega F(\omega)$
Parseval's theorem	$\int_{-\infty}^{\infty} \lvert f(t)\rvert^2 dt$	$\int_{-\infty}^{\infty} \lvert F(\omega)\rvert^2 \mathrm{d}\omega$
Auto-correlation	$f(t) * f^*(-t)$	$\lvert F(\omega)\rvert^2$

Table 2.1: *Fourier transform properties*

2.3.1 Probability

A signal that always occurs under a given set of conditions is called a certain or deterministic signal. Deterministic signals form an important basis for analysis in engineering, but the real world is inherently a much less ordered place. In practice most signals of interest, i.e., speech, biomedical, share market prices, etc. have a random component. In order to deal with these signals it is necessary to use statistics. Thus, probability theory is defined as the study of mathematical models of random phenomena [3].

It is convenient to define the sample space $\mathcal{S}$ as a set of all possible outcomes which can be either finite or infinite. An event is one particular outcome in the sample space. Assign to each event A a number $P_r\{A\}$ which is called the probability of the event A. This number is chosen so as to satisfy the following three conditions:

$$P_r\{A\} \geq 0 \tag{2.34}$$

$$P_r\{\mathcal{S}\} = 1 \tag{2.35}$$

$$P_r\{A+B\} = P_r\{A\} + P_r\{B\} \text{ if } AB = \{\emptyset\} \tag{2.36}$$

where the empty set $\emptyset$ is the impossible event. These conditions are the axioms of the

theory of probability where upon the development of the theory of probability is based on, directly or indirectly [3].

2.3.2 Random variables

In simple terms, a random variable is a function X that assigns a real number to each outcome x of an experiment [3]. Thus, the value x of a random variable X is not precisely predictable, i.e. one can only talk of the probability of it taking a particular value. Examples for discrete variables can be thought of as the value of the throw of a dice, or the number of photons detected. For an optical wavefront the random variable is continuous.

The *probability density function* (PDF), $f(x)$, is a valuable tool in analyzing probabilistic situations. It is the derivative of the *cumulative density function* (CDF) or the distribution function $F(x)$,

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}, \tag{2.37}$$

where

$$F(x) = P_r\{X \leq x\}. \tag{2.38}$$

There are a large number of probability distributions in common use. However, two specific PDFs illustrate the continuous and discrete cases; both are important in later work in this thesis.

Gaussian. A random variable X is called *Gaussian* or *normal* if its PDF has the following form

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \tag{2.39}$$

where μ is the mean and σ is the standard deviation. This distribution is widely used because it accurately represents a wide variety of physical situations and it has numerous convenient mathematical properties.

Poisson. A random variable X is *Poisson* distributed with a parameter μ as both the mean and variance if it takes the values $0, 1, \ldots, n, \ldots$ with

$$f_X(x) = \frac{\mu^m \exp[-\mu]}{m!} \qquad m = 0, 1, \cdots \tag{2.40}$$

The Poisson distribution is used to model a wide variety of physical processes such as photon (shot) noise, and detector readout noise, which are described in chapter 4.

2.3.3 Moments and covariance

Rather than specify the whole function $f(x)$, it is sometimes convenient to describe the behaviour of the random variable by a few numbers, the moments μ_n. For instance, the Gaussian distribution is completely specified by the first two moments. For other distributions more moments are needed to completely define the PDF.

The k^{th}-order moment of a random variable X, $k = 0, 1, 2, \cdots$ is defined by

$$\mu_k = \langle X^k \rangle = \int_{-\infty}^{\infty} x^k f_X(x) \mathrm{d}x \tag{2.41}$$

where $\langle\rangle$ is used to indicate the process of averaging over the ensemble (i.e. the collection of realizations). In many situations it is convenient to compute the moments about the mean value $\langle X \rangle$ rather than zero, which are called central moments,

$$\mu_{mk} = \langle (X - \langle X \rangle)^k \rangle = \int_{-\infty}^{\infty} (x - \langle X \rangle)^k f_X(x) \mathrm{d}x. \tag{2.42}$$

Of particular importance is the first moment or the *mean*,

$$\mu_1 = \langle X \rangle = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x, \tag{2.43}$$

and the second moment,

$$\mu_2 = \int_{-\infty}^{\infty} x^2 f_X(x) \mathrm{d}x. \tag{2.44}$$

The *variance* σ^2 is defined as

$$\sigma^2 = \mu_2 - \mu_1^2 \tag{2.45}$$

and the square root of the variance is called the *standard deviation.*

An extension to two joint random variables is useful for the work in this thesis, and it is now addressed.

Joint distribution A joint distribution occurs when an experiment has more than one outcome. For example, the problem of throwing an object on the floor. The outcome in this case can be described by two numbers representing the coordinates on the floor where the object lands. The joint (bivariate) distribution $F_{X,Y}(x, y)$ or simply, $F(x, y)$, of two random variables X and Y is written as

$$F(x, y) = P_r\{X \leq x, Y \leq y\} \tag{2.46}$$

and the joint density is by definition the function

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}, \tag{2.47}$$

where

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y = 1. \tag{2.48}$$

Independence Two random variables X and Y are called statistically independent if

$$P_r\{X,Y\} = P_r\{X\}P_r\{Y\}. \tag{2.49}$$

Hence, the CDF is

$$F(x,y) = F(x)F(y), \tag{2.50}$$

and the PDF is

$$f(x,y) = f(x)f(y). \tag{2.51}$$

Covariance The covariance C of two random variables X and Y is by definition:

$$\begin{aligned} C_{XY} &= \langle (X - \langle Y\rangle)(Y - \langle Y\rangle)\rangle \\ &= \int_{\infty}^{\infty}\int_{\infty}^{\infty} (x - \langle X\rangle)(y - \langle Y\rangle) f(x,y)\mathrm{d}x\mathrm{d}y. \end{aligned} \tag{2.52}$$

The covariance of a random variable X with itself is

$$\begin{aligned} C_{XX} &= \langle (X - \langle X\rangle)^2\rangle \\ &= \sigma_x^2. \end{aligned} \tag{2.53}$$

The correlation and auto-correlation functions are respectively defined similarly but without subtracting the mean value. The relationship between covariance C_{XY} and correlation R_{XY} can be expressed as

$$C_{XY} = R_{XY} - \langle X\rangle\langle Y\rangle. \tag{2.54}$$

Two random variables X and Y are uncorrelated if their covariance is 0. From Eq. (2.54) it can be seen that if two random variables are statistically independent, they are automatically uncorrelated. However, for two random variables which are uncorrelated, they are not necessarily statistically independent. This can be seen from the definition of correlation

$$R_{XY} = \langle XY\rangle = \langle X\rangle\langle Y\rangle \tag{2.55}$$

where R_{XY} does not have to be zero if X and Y are uncorrelated.

2.3.4 Conditional probability and Bayes rule

The probability of discrete events associated primarily with a single experiment has been discussed. However, in many applications an experiment is composed of sub-experiments. This involves the notion of conditional probability. Define the symbol $P_r\{A|B\}$ as representing the probability that event A occurs, conditioned upon the knowledge that B has already occurred. This is a conditional probability since only the outcomes are considered that include the event B and not all possible outcomes. The conditional probability of A given B is the number of times that both A and B occur divided by the number of times the event A occurs, and it may be written mathematically as [3]:

$$P_r\{A|B\} = \frac{P_r\{A,B\}}{P_r\{A\}}. \tag{2.56}$$

This expression is known as the conditional probability and may also be written

$$P_r\{B|A\} = \frac{P_r\{A,B\}}{P_r\{B\}}. \tag{2.57}$$

By combining Eqs. (2.56) and (2.57) such as to eliminate $P_r\{A,B\}$, it is straightforward to obtain a most useful expression known as *Bayes rule*

$$P_r\{A|B\} = \frac{P_r\{B|A\}P_r\{A\}}{P_r\{B\}}. \tag{2.58}$$

Bayes rule relates the two conditional probabilities $P_r\{A|B\}$ and $P_r\{B|A\}$ in terms of the marginal probabilities $P_r\{A\}$ and $P_r\{B\}$. $P_r\{A\}$ in Eq. (2.58) is often referred to as *a priori* information. Chapter 5 relies heavily on this rule since it allows use the *a priori* information to regularize a phase retrieval problem which is usually ill-posed (i.e. a small perturbation in the input causes a large deviation in the output).

2.3.5 Random processes

A *random process* $X(t)$ is a sequence of random variables in time or space, or both. For instance, the phase of a wave that has propagated through turbulence is a space and time varying random process. For explanation purposes, emphasis is placed here on functions of *time*. However, generalizations to functions of space are straightforward. The value of a random process at one point t_n is a random variable, and has a PDF $f_X(t_n)$. If the process is *stationary*, then none of these quantities depends on t, that is, the statistics are the same everywhere.

The collection of all realizations $X(t)$ of a stochastic process is called an *ensemble*. Figure 2.5 illustrates the ensemble of second order waveforms and a pair of parameter values t_1 and

t_2. The second order density function is the joint density function of the random variables $X(t_1)$ and $X(t_2)$. In general this density function depends on both t_1 and t_2 and hence is denoted $f(X_1, X_2; t_1, t_2)$, where $X_1 = x(t_1)$, $X_2 = x(t_2)$.

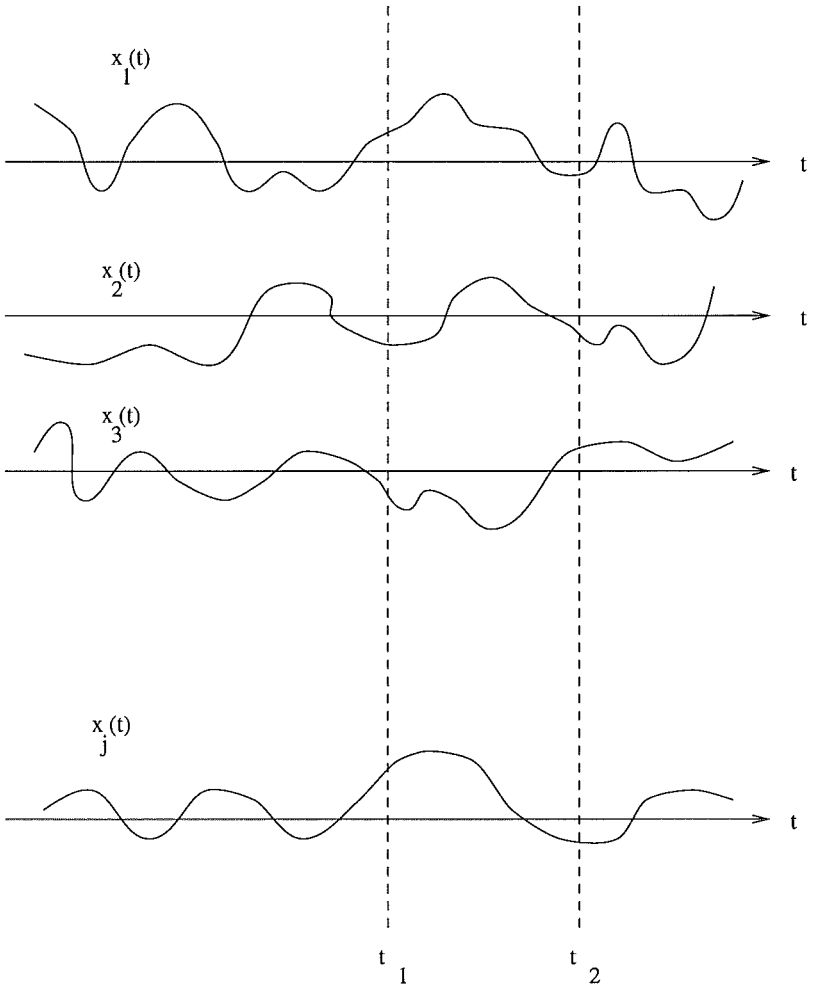


Figure 2.5: *An ensemble of sample functions, where t_1 and t_2 are the parameter values for which the joint density function $f(X_1, X_2; t_1, t_2)$ is specified.*

A random process is called *strictly stationary* if the j^{th} order joint probability density function $f(X_1, X_2, \ldots, X_j; t_1, t_2, \ldots, t_j)$ is independent of the choice of time origin for all j. In other words, the statistics are unchanged by shift of time. In many cases it is sufficient to have only both mean and autocorrelation function time invariant. This is called *wide*

sense stationary;

$$R_X(t_1, t_2) = R_X(t_2 - t_1). \tag{2.59}$$

A comparison of the properties of an individual sample function as it evolves along the time axis, with the properties of the entire ensemble at one or more specific instants of time, is referred to as *ergodicity*. A random process is called *ergodic* if every sample function takes on values along the time axis (horizontally) with the same joint relative frequencies observed across the ensemble at any instant or collection of instants (vertically). For a random process to be ergodic, it is necessary that it be strictly stationary. The hierarchy of classes of random processes is depicted in Fig. 2.6.

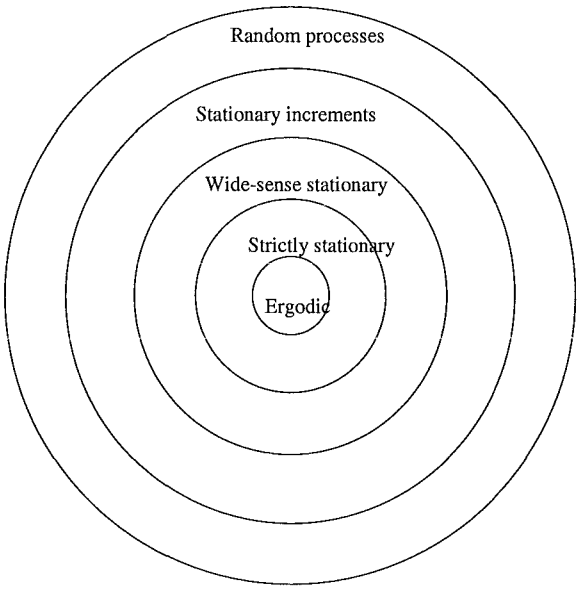


Figure 2.6: *The hierarchy of classes of random processes.*

A function $f(x)$ is strictly called the *first order* PDF since it only describes the statistics at a single point x. In the case of spatial or temporal structure of a random process, for example is it changing slowly or rapidly, it is necessary to use a *second order* PDF. Similarly, the second order PDF can be parameterized in terms of its moments. By far the most important moment is the covariance function $C(x')$ which can be written as follows for a stationary real process:

$$C(t') = \langle (X(t) - \langle X \rangle)(X(t + t') - \langle X \rangle) \rangle. \tag{2.60}$$

Note that often $\langle X \rangle = 0$ and that the variance is given by

$$\sigma^2 = C(t' = 0) = \langle X^2 \rangle - \langle X \rangle^2. \tag{2.61}$$

In the case of atmospheric turbulence presented in the next chapter, it is sometimes more convenient to work with a *structure function* which is the mean square difference of two random processes as defined by

$$\begin{aligned} \mathcal{D}(t') &= \langle (X(t) - X(t+t'))^2 \rangle \\ &= 2[C(0) - C(t')]. \end{aligned} \tag{2.62}$$

From Eq. (2.62) it can be seen that the structure function has a value of zero for zero separation, and only a small value for small separations. Using the knowledge that the power spectrum and the autocorrelation function have a Fourier transform relationship (table 2.1),

$$\Phi(\kappa) = \int_{-\infty}^{\infty} C(t') \exp[-2\pi i \kappa t'] \mathrm{d}t', \tag{2.63}$$

it is possible to find the relation between the structure function and the power spectrum as

$$\begin{aligned} \mathcal{D}(t') &= 2 \int \Phi(\kappa) \left[1 - \exp[2\pi i \kappa t']\right] d\kappa \\ &= 2 \int \Phi(\kappa) \left[1 - \cos(2\pi i \kappa t')\right] d\kappa. \end{aligned} \tag{2.64}$$

The structure function is in some cases preferred to the covariance since it provides a degree of information to the points in a finite aperture. It is impossible to define a covariance function for points in a finite aperture since the covariance of the DC term is infinite. It is only possible to define the covariance over a finite region when the mean is removed [9,38].

Chapter 3

REVIEW OF IMAGING THROUGH TURBULENCE

This chapter reviews relevant theory regarding the optical effects of the atmospheric turbulence. This review provides a basis for the new wavefront estimation techniques introduced in chapters 4 and 5. This chapter is structured in the following manner.

In section 3.1 the fundamentals of optical imaging are presented. A wavefront as a key concept in optics is introduced here. It is also addressed how the Fourier transform discussed in the previous chapter closely models the relationship between the source of wave motion and the propagated field at a distance. This helps explain the wide applicability of the Fourier transform throughout science and engineering. Further the concept of coherent and incoherent imaging is explained.

The properties of atmospheric turbulence are described in section 3.2. The Kolmogorov model, a common model for describing the turbulence, is explained. A random screen model for the turbulence is then presented through computer simulation. The random screen is placed in the pupil plane of the imaging system and effectively models the atmospheric induced field perturbations. The short and long exposure which refer to the exposure time are finally addressed.

Section 3.3 concludes with a brief introduction to the latest technology in adaptive optics

systems, namely the laser guide star. The main problems of this technique associated with cone effect and angular isoplanatism are briefly discussed. Furthermore, in section 3.3.2 presents a notion of relating the adaptive optics system to phase retrieval.

3.1 Fundamentals of Optical Imaging

The aim of this section is to provide a strong basic introduction to the next two chapters. The primary motivation here is to develop an understanding of the way optical systems process light to form images.

3.1.1 Optical path and wavefront aberrations

In this section, the principles of geometrical optics are discussed. It is assumed that light travels in straight lines and is partially refracted and reflected at an interface.

In order to discuss optical path, it is necessary to introduce the *refractive index* which is the value of the constant ratio n_{12} for refraction from the first into the second medium. The refractive index for refraction from vacuum into a medium, is called the *absolute refractive index* and it is defined as,

$$n = \frac{c}{v} = \sqrt{\epsilon\mu} \tag{3.1}$$

where c and v are the speed of light in vacuum and in the medium, respectively, ϵ is known as the dielectric constant or permitivity, and μ is called the magnetic permeability [30].

According to Snell's law [31], if a plane electro-magnetic wave falls on to a boundary between two homogeneous media, the sine of the angle θ_1 between the normal to the incident wave and the normal to the surface bears a constant ratio to the sine of the angle θ_2 between the normal of the refracted wave and the surface normal (Fig. 3.1). This constant ratio is equal to the ratio of the velocities v_1 and v_2 of propagation in the two media [31]:

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2} \tag{3.2}$$

The laws of reflection and refraction, and the manner in which light propagates in general, can be viewed from the Fermat principle. This principle states that the actual path between two points taken by a beam of light is the one that is traversed in the least time [30]. This can be seen rather easily in Fig. 3.1, which depicts a point source A emitting a ray that

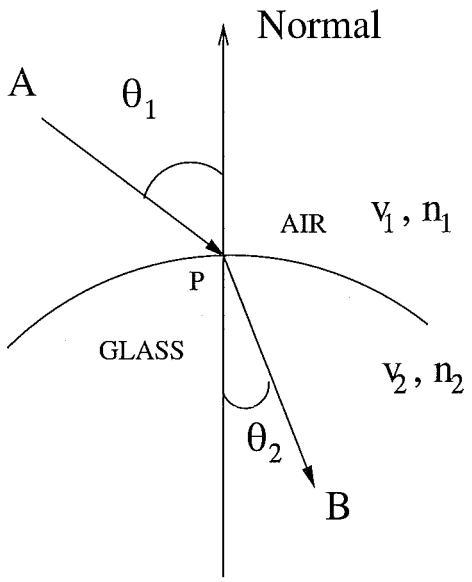


Figure 3.1: *Fermat's principle applied to refraction.*

is then refracted toward B. The transit length from A to B is known as the *optical path length* and can be stated in the form [30, 31]

$$[AB] = \int_A^B n \, \mathrm{d}s, \tag{3.3}$$

where $\mathrm{d}s$ is a differential element of length along any one of the paths from A to B. Thus the optical path from a point source to all points on a given wavefront is equal to a constant which is the product of the actual path and the refractive index of the medium of which the light travels through.

Another fundamental model of light useful in optics is the *wavefront*. A wavefront is found by tracing out an equal distance along the optical path from a source to the region of interest, for example, the entrance pupil of an optical system. A spherical wavefront of a point source is illustrated in Fig. 3.2. The importance of a wavefront is further discussed in detail in chapter 4 which explains how to sense a wavefront before any corrections can be made.

An ideal optical system simply means that a point source in object space also becomes a point source in image space (Fig. 3.3). However, a real optical system has aberrations. There are two types of aberrations, one is point imaging aberrations where the rays in image space are not concurrent, and the other one is aberrations of image shape, in which

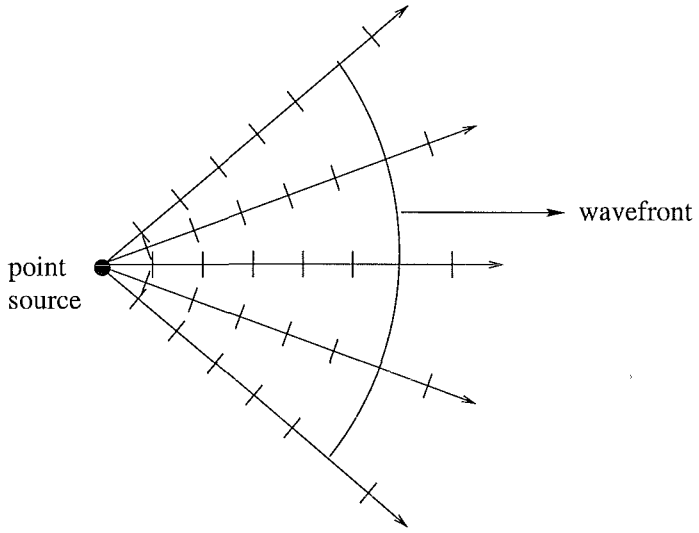


Figure 3.2: *Wavefront is found by tracing out optical paths with equal distance from the source.*

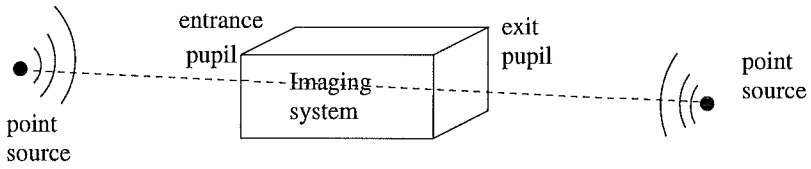


Figure 3.3: *Ideal optical system.*

each object point forms a true image point but there is not the correct similarity between object and image shapes. The latter aberrations are due to field curvature and distortion while the former are spherical aberration, coma, astigmatism, and higher order compounds of complex form [30].

Having defined wavefronts and aberrations, a combination of these two is now explained. Let S' be the wavefront of a point source which passes through O and let S be a *reference sphere* with center P_0 (see Fig. 3.4). Let another ray r of the point source meet S and S' in Q_0 and Q respectively, and let it meet the image plane at P. The wavefront aberration, W, is then defined as the optical path length from Q_0 to Q, i.e. nQ_0Q, where n is the refractive index of the space in between. Clearly the function $W(x, y)$ expresses the deformation of the wavefront from the ideal spherical shape. It is useful to carry out an expansion in terms of the wavefront aberration W. In polar coordinates $W(x, y)$ can be written as $W(\rho, \theta)$, where $\rho = \sqrt{x^2 + y^2}$, $\theta = \tan^{-1}(y/x)$, and it can be expanded as a power series in these

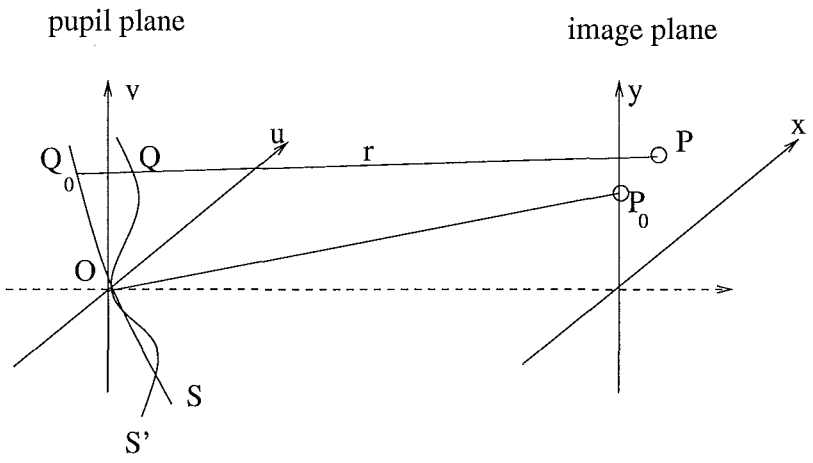


Figure 3.4: *Wavefront aberration defined as the optical path length from Q_0 to Q.*

variables as [31]

$$\begin{aligned} W(\rho^2, \rho y\cos\theta, y^2) &= a_1\rho^2 + a_2\rho y\cos\theta \\ &+ b_1\rho^4 && \text{spherical aberration} \\ &+ b_2\rho^3 y\cos\theta && \text{coma} \\ &+ b_3\rho^2 y^2\cos\theta && \text{astigmatition} \\ &+ b_4\rho^2 y^2 && \text{field curvature} \\ &+ b_5\rho y^3\cos\theta && \text{distortion} \end{aligned} \tag{3.4}$$

The last five terms in Eq. (3.4) are called the *Seidel* aberrations [30]. An example of spherical aberration of form $b_1\rho^4$ occurred recently in the Hubble Space Telescope. It is important to note here that when $W(x, y) = 0$, the ultimate performance of the imaging system is reached. This case is often referred to as diffraction-limited imaging.

3.1.2 Fourier optics

Obviously, a real optical system cannot give perfect images, even when there are no aberrations. This is due to the diffraction of the optical system and can be explained through the following diffraction theory.

Consider a generalized imaging system shown in Fig. 3.5. The aperture (e.g. telescope pupil) is in the (u, v)-plane and the image plane has coordinates (ζ, η). The wave motion is characterized by a complex-valued scalar function of position and time. In this thesis

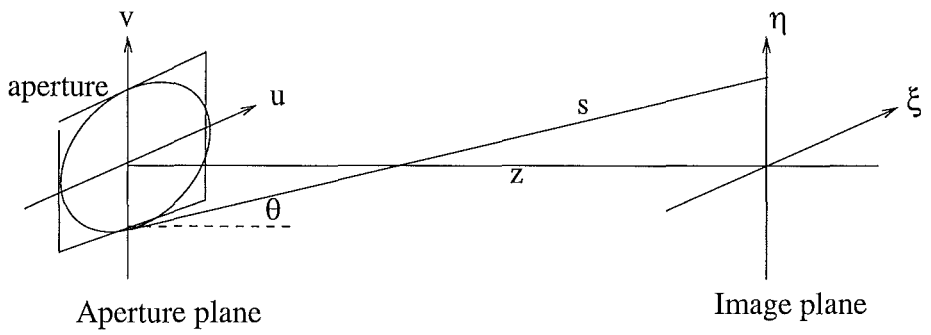


Figure 3.5: *Geometry used in explaining the diffraction theory.*

the wave function is generalized to account for both the finite pupil size and pupil plane aberrations. The generalized pupil function may be interpreted as the product of a physical pupil function and a function describing the aberration [34]:

$$P(x,y) = \begin{cases} P_a(x,y)\exp[i\phi(x,y)] & \text{inside aperture} \\ 0 & \text{outside aperture} \end{cases} \tag{3.5}$$

where $P_a(x,y)$ is a real function describing any amplitude modulation due to the aberrations, and the wavefront phase, $\phi(x,y)$, is related to the wavefront function itself by:

$$\phi(x,y) = \frac{2\pi}{\lambda} W(x,y). \tag{3.6}$$

The parameter $2\pi/\lambda$ is often referred to as the wave number k.

It is convenient to define z to denote the distance between the aperture plane and the image plane. The fields in the image and aperture planes are denoted $h(\zeta,\eta)$ and $P(\lambda z u, \lambda z v)$, respectively. Having defined the image and aperture planes it is possible to express the diffracted field in terms of the aperture field by use of the *Rayleigh-Sommerfeld* formula [34]

$$h(\zeta,\eta) = \frac{1}{i\lambda}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(\lambda z u, \lambda z v)\frac{\exp[iks]}{s}\cos(\theta)\mathrm{d}u\mathrm{d}v \tag{3.7}$$

where s is a distance shown in Fig. 3.5 joining an arbitrary point in the aperture plane to an arbitrary point in the image plane, and $\cos(\theta)$ is the cosine of the angle between s and the z-axis. The term $\exp[iks]/s$ in Eq. (3.7) represents a unit amplitude spherical wave expanding about the point (u,v). The formula in Eq. (3.7) shows that the field at any point behind the aperture is the same as that which would be produced by a secondary source at the aperture.

If it is assumed that z is much greater than the aperture dimension and λ, the factor $\cos(\theta)$ in Eq. (3.7) may be replaced by unity. Under these conditions, the factor $1/s$ may be

replaced by $1/z$. The s term in the exponent may not be replaced with z, however, since $2\pi/\lambda$ is large in many situations which would amplify the phase error introduced by the approximation. The exact expression for s is

$$s = \sqrt{z^2 + (u-\zeta)^2 + (v-\eta)^2}, \tag{3.8}$$

which can be expanded by use of the binomial theorem. Discarding terms of higher than first order yields

$$s \approx z + \frac{(u-\zeta)^2}{2} + \frac{(v-\eta)^2}{2} \tag{3.9}$$

which when substituted in Eq. (3.7), yields the *Fresnel* or *near field* approximation [34]:

$$h(\zeta,\eta) = K \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(\lambda z u, \lambda z v)\exp[ik\frac{(u^2+v^2)}{2z}]\exp[ik\frac{1}{z}(u\zeta + v\eta)]\mathrm{d}u\mathrm{d}v \tag{3.10}$$

where K is a complex constant.

For most astronomical applications it is reasonable to assume that the source is very distant so that the diffracting aperture is illuminated by plane wavefronts. Reformulating the Fresnel theorem; a plane wavefront entering the aperture generates new spherical wavefronts. When z is further increased, the waves arriving at the image plane are again plane, and therefore the quadratic terms of the Fresnel transform can be discarded. This case is often referred to as *Fraunhofer* or *far field* approximation, and the integral in Eq. (3.10) results in [34]

$$h(\zeta,\eta) = K \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(\lambda z u, \lambda z v)\exp[ik\frac{1}{z}(u\zeta + v\eta)]\mathrm{d}u\mathrm{d}v \tag{3.11}$$

which represents simply a Fourier transform relationship between the aperture and image planes (section 2.2). Fortunately, the simpler case of Fraunhofer diffraction suffices in most situations.

Consider Fraunhofer diffraction at a circular aperture and in the absence of aberration, that is when $W(x,y) = 0$, Eq. (3.11) becomes

$$h(\zeta,\eta) = K \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp[ik\frac{1}{z}(u\zeta + v\eta)]\mathrm{d}u\mathrm{d}v. \tag{3.12}$$

Using the Hankel transform described in section 2.2.5, the diffraction integral in Eq. (3.12) results in a form of the first order Bessel function [34]:

$$h(\zeta,\eta) = \frac{2J_1(Z)}{Z}, \tag{3.13}$$

where

$$Z = \frac{2\pi}{\lambda}\frac{a}{z}\sqrt{\zeta^2+\eta^2}, \tag{3.14}$$

and a is the the radius of the circular aperture. Hence the intensity is given by

$$|h(\zeta,\eta)|^2 = \left(\frac{2J_1(Z)}{Z}\right)^2, \tag{3.15}$$

which is well-known as the Airy function, and is depicted in Fig. 3.6. It is important to note

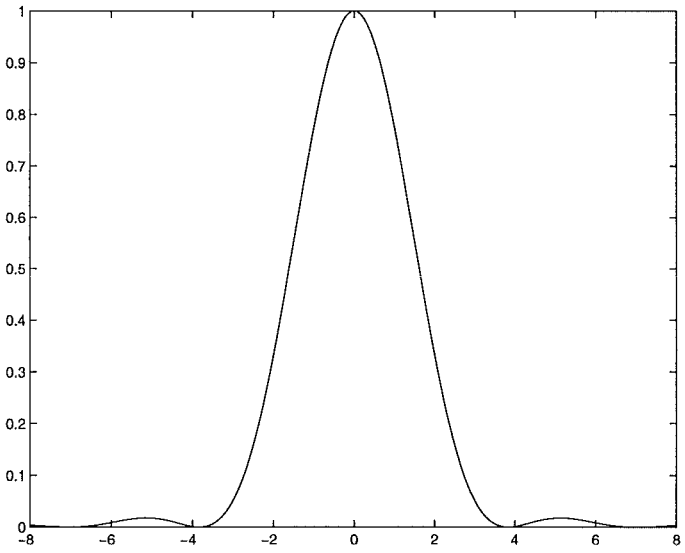


Figure 3.6: *One-dimensional Airy function.*

here that the function possesses side-lobes after its first minimum, and that the minima of the function are not equidistant as shown in table 3.1.

Z	$\left(\frac{2J_1(Z)}{Z}\right)^2$	
0	1	Max
3.833	0	Min
5.136	0.0175	Max
7.016	0	Min
8.417	0.0042	Max
10.174	0	Min

Table 3.1: *The first few maxima minima of the function* $|h(\zeta,\eta)|^2 = \left(\frac{2J_1(Z)}{Z}\right)^2$

Now consider the resolution of the diffraction-limited imaging. The resolving power of a

telescope may be defined as the ability of a telescope to separate objects which subtend small angles to the observer [32]. Even if it were possible to remove all the aberrations from the telescope optical system, there still would be fundamental limit to the ability of any telescope to separate objects which are close together. If two point sources are very close together, the resulting image is the superposition of two diffraction patterns. It is only possible to resolve the resultant image as being made of two components if the individual Airy disks are sufficiently separated. According to *Rayleigh*'s criterion for resolution, the two images are said to be resolvable when the center of one Airy disk falls on the first minimum of the other diffraction pattern, corresponding to an approximately 20% drop in intensity, a point illustrated in Fig. 3.7.

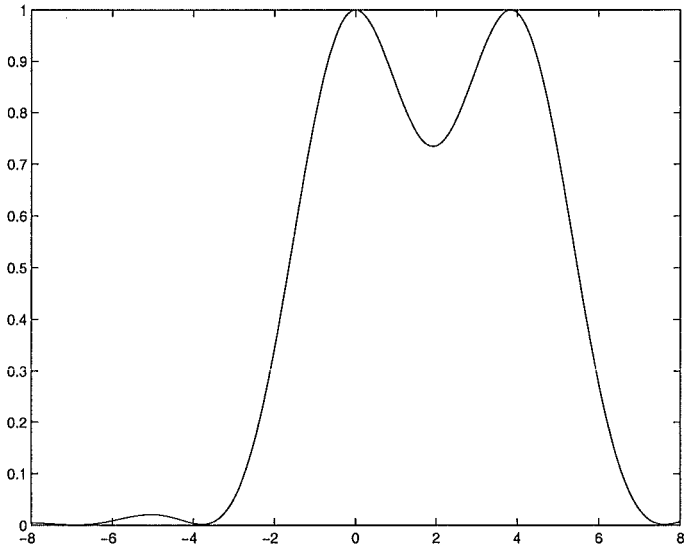


Figure 3.7: *The Rayleigh resolution.*

Using Eq. (3.14) and the first minimum listed in Table 1, the minimum radius of the Airy disk allowing two images to be resolved can be calculated as

$$r_{min} = 1.22\lambda F \quad [\text{m}], \tag{3.16}$$

where F is the F-number defined as $F = z/2a$ [31]. In stellar astronomy, however, the separation between objects is commonly specified in angular units rather than in linear measure. In terms of angles, it should thus be possible to resolve two stars if they are separated by an angle greater than

$$\alpha_{min} = 1.22\frac{\lambda}{D} \quad [\text{rad}] \tag{3.17}$$

where D is the aperture dimension.

For relatively weak aberrations the main effects of wavefront errors are a broadening of the PSF and a consequent reduction of its peak value. Under these circumstances a useful figure of merit is the Strehl intensity or Strehl ratio, S. This is defined as the ratio of the irradiance of the peak of the PSF and the irradiance obtained in the diffraction-limited case [4],

$$S = \frac{(|h(x,y)|^2_{max})_{\text{actual}}}{(|h(x,y)|^2_{max})_{\text{diffraction limited}}} \qquad 0 \leq S \leq 1. \tag{3.18}$$

This ratio is illustrated in Fig. 3.8 as P_1/P_0, where P_0 is the peak of the diffraction-limited PSF and P_1 is that of an aberrated system.

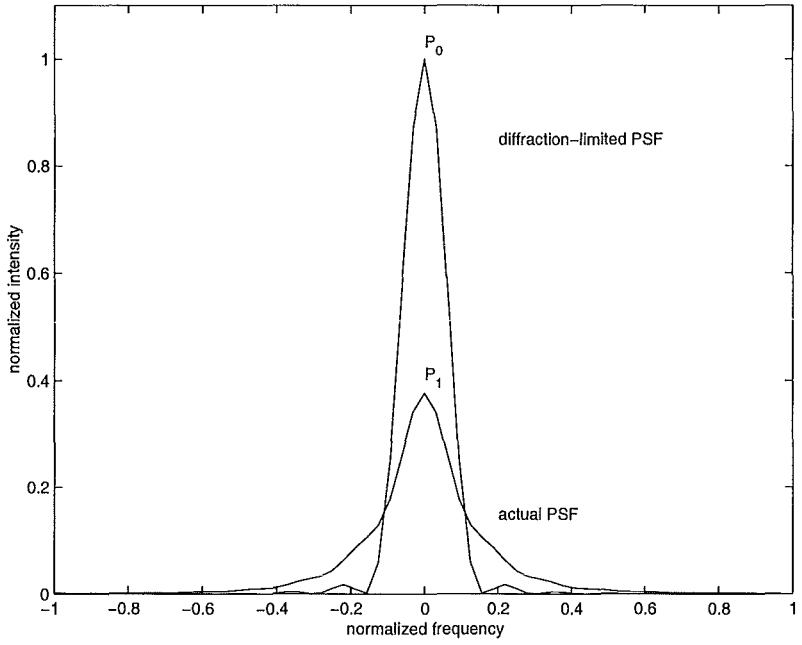


Figure 3.8: *Point spread function illustrating the Strehl ratio, P_1/P_0.*

In the weak aberration limit, a useful result is that the focal plane irradiance can be expressed as a function of the mean square wavefront error [1],

$$S = \frac{1}{\pi^2}\left|\int_0^1 \int_0^{2\pi} \exp[-i\frac{2\pi}{\lambda}W(\rho,\theta)]\rho \, \mathrm{d}\rho\mathrm{d}\theta\right|^2. \tag{3.19}$$

Note that if there are no aberrations or $W(\rho,\theta) = 0$, S becomes unity. For small aberrations,

that is when

$$\exp[-i\frac{2\pi}{\lambda}W] \approx 1 - i\frac{2\pi}{\lambda}W - \frac{1}{2}\left(i\frac{2\pi}{\lambda}\right)^2 W^2 + \ldots, \tag{3.20}$$

the Strehl ratio is related to the variance of the phase aberration, σ_ϕ^2, by

$$S \approx 1 - \sigma_\phi^2 \qquad \sigma_\phi << 1. \tag{3.21}$$

A common convention is to identify a diffraction-limited imaging capability with a Strehl ratio ≥ 0.8, that is when $\sigma_\phi^2 \leq 0.2$. This condition is often referred to as *Marechal*'s criterion.

If the phase aberration is not small, then the Strehl ratio depends on the specific form of the aberration. However, for a random Gaussian aberration, such as that given by atmospheric turbulence, the Strehl can be expressed as

$$S = \exp[-\sigma_\phi^2] \tag{3.22}$$

which is a property of Gaussian random processes.

3.1.3 Coherent and incoherent imaging

In order to describe coherent and incoherent imaging, it is necessary to have a complete understanding about thermal and laser light. Many optical sources emit light by means of spontaneous emission from a collection of molecules. The process of this light emission occurs when a large collection of atoms or molecules randomly and independently drop from high energy states to lower energy states. Such radiation, consisting of a large number of independent contributions, is referred to as *thermal light* [33]. In contrast, the relatively well-ordered stimulated radiation emitted by a laser is referred to as *laser light* [33].

Thus, *coherent imaging* is when the object is illuminated by laser light, whereas *incoherent imaging* is referred to as an object which is illuminated by thermal light or is self-luminant. Considering this concept from the types of object illumination, coherent illumination is when all point sources in the plane have a fixed phase relationship or the optical field is perfectly correlated. Object illumination with the opposite property, that is when the point sources vary in phase independently of each other, is referred to as incoherent illumination. It is important to note that coherent illumination is linear in complex amplitude, whereas incoherent illumination is linear in intensity [33].

Consider first the coherent imaging. Assuming the impulse response (PSF) is space-invariant, the relationship between the input and output of the complex amplitude of the

image can be modeled as a convolution of the input image and the PSF,

$$U_{out}(\zeta,\eta) = U_{in}(\zeta,\eta) \odot h(\zeta,\eta), \tag{3.23}$$

where $U(\zeta,\eta)$ refers to the complex amplitude of either the input or the output image. As described in section 2.2, a convolution in the image space is equivalent to a multiplication in the Fourier space, and hence Eq. (3.23) becomes

$$\tilde{U}_{out}(u,v) = \tilde{U}_{in}(u,v)H(u,v), \tag{3.24}$$

where a symbol $\tilde{}$ above a function denotes the Fourier transform of the function, and $H(u,v)$ is the Fourier transform of $h(\zeta,\eta)$. Recalling Eq. (3.11), it can be deduced that,

$$\begin{aligned} H(u,v) &= \mathcal{F}\{\mathcal{F}\{P(-\lambda z u, -\lambda z v)\}\} \\ &= P(-\lambda z u, -\lambda z v) \end{aligned} \tag{3.25}$$

which is referred to as the coherent transfer function. Hence, the coherent transfer function is proportional to the reflected pupil function.

Consider now the incoherent imaging. As incoherent illumination is linear in intensity, the relationship between the input and output of the complex amplitude of the image can be expressed as

$$U_{out} = U_{in} \odot |h(\zeta,\eta)|^2. \tag{3.26}$$

In Fourier space it becomes,

$$\tilde{U}_{out}(u,v) = \tilde{U}_{in}(u,v)H(u,v), \tag{3.27}$$

where

$$\begin{aligned} H(u,v) &= \int\int |h(\zeta,\eta)|^2 \exp[-2\pi i(u\zeta + v\eta)]\mathrm{d}\zeta\mathrm{d}\eta \\ &= \int\int P^*(\lambda z u', \lambda z v')P(\lambda z u' + \lambda z u, \lambda z v' + \lambda z v)\mathrm{d}u'\mathrm{d}v' \\ &= P * P. \end{aligned} \tag{3.28}$$

Therefore, the transfer function of an incoherent optical system is proportional to the autocorrelation of the pupil function.

3.2 Properties of Atmospheric Turbulence

Having discussed the wavefront aberration and the optical imaging system, consider now the main cause of the aberrations, which is the atmospheric turbulence. Atmospheric turbulence

affects all optical systems which use light that has propagated through long atmospheric paths. These effects limit our ability to measure high angular resolution information about the objects of interest.

This section deals mainly with the properties of this turbulence, models for characterizing the strength of the turbulence, techniques for generating phase screens to simulate the effects of the turbulence, and the short and long exposure imaging.

3.2.1 Kolmogorov turbulence

The knowledge of the statistics of atmospheric turbulence comes from dimensional analysis and simple physical arguments. The key result concerns the statistics of turbulent velocity and is due to Kolmogorov [41]. Tatarski extended the statistics of turbulent velocity results to atmospheric refractive index variations [4].

The Kolmogorov result was based on the mean-square-difference of velocity $v(\mathbf{r})$, measured at two points, $\mathbf{r}_1$ and $\mathbf{r}_2$, i.e. the velocity structure function, $\mathcal{D}_v(\mathbf{r}_1, \mathbf{r}_2)$ discussed in section 2.3.5, where $\mathbf{r}$ is a three-dimensional vector. He made a hypothesis that this structure function should be related to the absolute distance between the two points expressing

$$\mathcal{D}_v(\mathbf{r}_1, \mathbf{r}_2) \propto |\mathbf{r}_1 - \mathbf{r}_2|^{2/3}. \tag{3.29}$$

Tatarski's contribution here was to introduce the concept that within a range of separations $\Delta\mathbf{r} = |\mathbf{r}_1 - \mathbf{r}_2|$ that are greater than the inner scale ℓ_0 and less than the outer scale L_0 (Fig. 3.9), the turbulence is the result of non-uniform temperature fluctuations at any altitude,

$$\mathcal{D}_T = C_T^2 \Delta\mathbf{r}^{2/3} \qquad \ell_0 < \Delta\mathbf{r} < L_0 \tag{3.30}$$

where C_T^2 is a temperature structure constant which has to be determined experimentally. In terms of the refractive index fluctuation, Eq. (3.30) can be written as,

$$\begin{aligned} \mathcal{D}_n(\mathbf{r}_1, \mathbf{r}_2) &= \langle (n(\mathbf{r}_1) - n(\mathbf{r}_2))^2 \rangle \\ &= C_N^2 \Delta\mathbf{r}^{2/3} \end{aligned} \tag{3.31}$$

where $n(\mathbf{r})$ is the refractive index at position $\mathbf{r}$, and the quantity C_N^2 is referred to as the refractive index structure constant which is also to be determined experimentally and has the units of $\mathrm{m}^{-2/3}$.

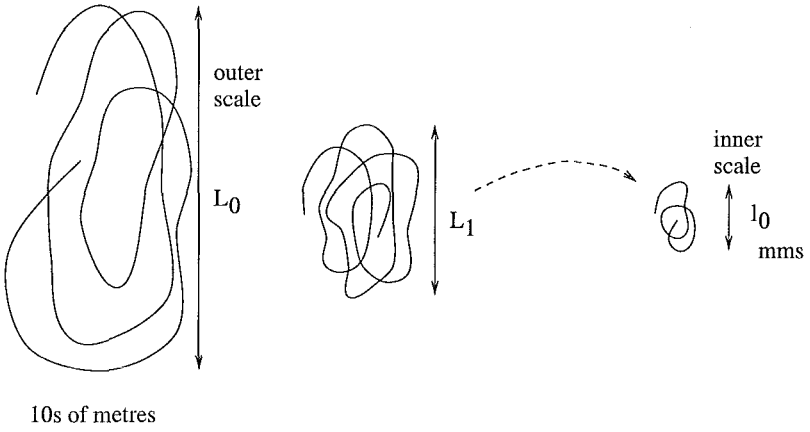


Figure 3.9: *The outer scale (L_0) and inner scale (ℓ_0) of the Kolmogorov turbulence.*

3.2.2 Field correlation and phase structure function

For convenience consider only horizontal monochromatic plane waves, of wavelength λ, propagating downward from a star towards a ground based observer. The complex amplitude $U_h(\mathbf{r})$ of a plane wave that has propagated through turbulence is given by

$$U_h(\mathbf{r}) = |U_h(\mathbf{r})|\exp[i\phi_h(\mathbf{r})] = |U_h(\mathbf{r})|\exp[i\frac{2\pi}{\lambda}W(\mathbf{r})], \tag{3.32}$$

where $\phi_h(\mathbf{r})$ is referred to its average value so that for any height h, $\langle\phi_h(\mathbf{r})\rangle = 0$. Furthermore, it is common to consider that the statistical properties of the atmosphere depend only upon the altitude h, a point illustrated in Fig. 3.10.

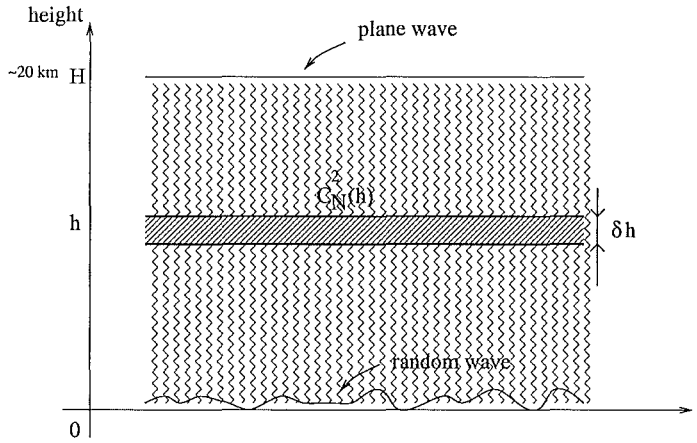


Figure 3.10: *A plane wave that is propagated through turbulent layers.*

A detailed analysis and derivation of the properties of the complex amplitude $U_h(\mathbf{r})$ can be found in [1, 55], and the results are simply quoted here. Roddier [55] shows that the covariance function of the complex amplitude for a wave that has propagated through Kolmogorov turbulence is given by

$$\begin{aligned} C_h(\mathbf{r}\,') &= \langle U_h(\mathbf{r})U_h^*(\mathbf{r}+\mathbf{r}\,')\rangle \\ &= \langle \exp[i\{\phi(\mathbf{r})-\phi(\mathbf{r}+\mathbf{r}\,')\}]\rangle \\ &= \exp[-\frac{1}{2}\mathcal{D}_\phi(\mathbf{r}\,')] \end{aligned} \tag{3.33}$$

where $\mathcal{D}_\phi(\mathbf{r}\,')$ is the phase structure function. This in turn is given by [55]

$$\mathcal{D}_\phi(\mathbf{r}\,') = 2.91k^2(\cos\gamma)^{-1}|\mathbf{r}\,'|^{5/3}\int_0^\infty C_N^2(z)\mathrm{d}z \tag{3.34}$$

$$= 6.88\left(\frac{|\mathbf{r}\,'|}{r_0}\right)^{5/3} \tag{3.35}$$

where r_0, the Fried parameter, is given by [41]

$$r_0 = \left[\frac{2.91}{6.88}k^2(\cos\gamma)^{-1}\int_0^H C_N^2(h)\mathrm{d}h\right]^{-3/5}. \tag{3.36}$$

This shows that r_0 depends on the zenith angle γ as $(\cos\gamma)^{-1}$ and on the wavelength λ as $\lambda^{6/5}$.

It can be seen from Eqs. (3.33) and (3.36) that the Fried parameter r_0 is a crucial quantity, since if this is known then both the phase structure function and field correlation can also be determined. This parameter also describes the aperture which has the same resolution as a diffraction-limited aperture in the absence of the turbulence. A typical value for r_0 in the visible is 10 cm; this means therefore that the area over this 10 cm aperture is equivalent to that of a diffraction-limited optical system of 10 cm aperture. In other words, if $r_0 = 10$ cm for a 10 m telescope, the image resolution is no better than that given by a 10 cm telescope.

As the wavelength increases, as mentioned earlier, for example in the infrared region of the spectrum, the size of r_0 also increases, and therefore the seeing improves. It should be noted, however, that the wavefront aberration introduced by turbulence is achromatic, i.e. independent of wavelength of light, so that, for example, the number of microns of optical path retardance is the same for both visible and infrared light. However, the number of wavelengths of optical path is different, i.e. there are fewer waves of retardance at (the longer) infrared wavelengths, and therefore, from this elementary picture, it is clear that the distortion experienced by infrared light is less serious than that experienced by visible light. It immediately follows that infrared wavelengths should be easier to compensate for.

3.2.3 Simulation of Kolmogorov turbulence

This section presents simulation approaches of atmospherically distorted wavefronts. Simulations are widely used to predict the performance of adaptive optics systems and various post processing schemes for imaging through turbulence. Many of the practical problems associated with the analytical computations are solved with the use of simulations. Moreover simulations can be used to provide images which show the effects of all noises and system limitations present in the optical system.

The starting point for nearly all analysis of atmospheric turbulence has been the assumption that atmospheric turbulence follows a Kolmogorov spectrum [55]. The fluctuations induced by the turbulence then cause a distortion of both the magnitude and phase of the wavefront incident on the atmosphere. In practice the quality of the images formed by light passing through the turbulence is considerably more affected by phase distortion than by magnitude distortion. In the case of static turbulence, an adequate approximation is often a single phase screen located at the entrance pupil of the optical system. By use of the Fourier transform presented in the previous chapter, it is then possible to propagate the simulated phase screen through the imaging system to form degrading speckle images.

There are a number of approaches used to generate random phase screens with the proper spatial and temporal correlation properties. Of these approaches the one that offers the simplicity is a direct approach based on the Kolmogorov model [37]. The first simulation approach presented here is implemented using this method. Ideal Kolmogorov turbulence is both infinite in extent and infinite in detail, and it is fractal in nature [36]. Given the two-dimensional phase structure function in Eq. (3.34) and recalling Eq. (2.64), the power spectrum of the phase fluctuation can be computed as:

$$\Phi_\phi(\mathbf{k}) = \frac{0.023}{r_0^{\frac{5}{3}}} |\mathbf{k}|^{\frac{-11}{3}}. \tag{3.37}$$

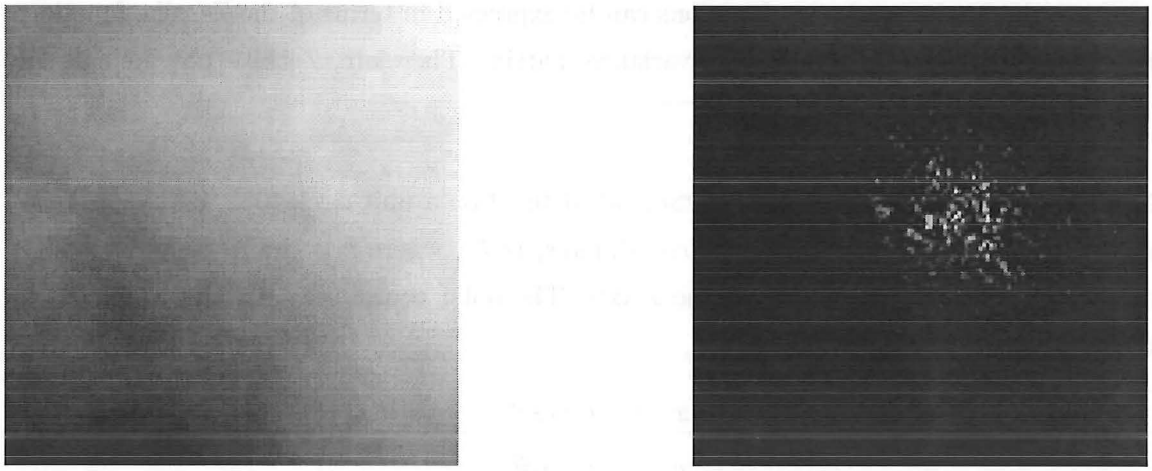
This expression for $\Phi(\mathbf{k})$ is also required when simulating Kolmogorov turbulence using McGlamery's method [35]. One may wish to add an outer scale term:

$$\Phi_\phi(\mathbf{k}) = \frac{0.023}{r_0^{\frac{5}{3}} \left(|\mathbf{k}|^2 + K_0^2\right)^{\frac{11}{6}}} \tag{3.38}$$

where $K_0 = 2\pi/L_0$, to avoid singularity as $|\mathbf{k}| \to 0$.

Figure 3.11a shows a typical two-dimensional phase screen obtained using this approximation, and the corresponding speckle image with $D/r_0 = 10$ is shown in Fig. 3.11b. A square

aperture is chosen, which is both mathematically and computationally convenient. The outer scale is set to $L_0 = 100$ m. For the simulation, the length of one side of the square aperture is set equal to 1 m, and there are 64 phase sample points across one side of the aperture, which provides a speckle image of size 128 × 128 samples to meet the Nyquist criterion (section 2.2.7.



(a) (b)

Figure 3.11: *(a) A typical phase screen produced by sampling the Kolmogorov spectrum directly (b) The corresponding speckle image ($D/r_0 = 10$).*

The second approach is based on a basis function represented by polynomials. Fried [41] found that the optical effects of air turbulence, which at first appear complex and random, can be described in terms of simple wavefront aberrations such as tip, tilt, defocus, and astigmatism (section 3.1), which are familiar to all workers in optics. These shapes are called the basis functions in mathematical terminology. Therefore an incident wavefront onto the telescope aperture can be written in terms of a set of orthogonal basis functions as,

$$\phi(x,y) = \sum_{j}^{\infty} a_j \psi_j(x,y) \tag{3.39}$$

where $\phi(x,y)$ represents the unknown wavefront incident onto the telescope aperture, $\psi_j(x,y)$ is the j^{th} basis function, and a_j is the weighting attached to the j^{th} basis function. This simulation approach is usually used when estimating wavefronts using a wavefront sensor which is discussed in the next chapter.

The optimal set of basis functions to describe a wavefront aberrated by Kolmogorov turbulence are the Karhunen-Loeve functions [39]. Choosing $\psi_j(x,y)$ to be Karhunen-Loeve functions means the coefficients a_j in Eq. (3.39) are statistically independent. Unfortunately, it is not possible to calculate these functions analytically. Noll [38] originally used Zernike polynomials to describe the atmospheric turbulence, although he showed that the Zernike coefficients are weakly correlated. However, Roddier [39] showed that for a circular aperture the Karhunen-Loeve functions can be expressed in terms of the Zernike functions by diagonalization of the Zernike covariance matrix. Therefore, Zernike polynomials are often used as the basis functions.

Zernike polynomials are a set of polynomials defined on a unit circle [38]. Zernike polynomials are generally written in polar coordinates, (r,θ), where r is the radius, $0 \leq r < 1$, and θ is the angle with respect to the x-axis. The polar coordinates, (r,θ), are related to the Cartesian coordinates, (x,y), by

$$\begin{aligned} x &= r\,\cos\theta, \\ y &= r\,\sin\theta, \\ r &= \sqrt{x^2+y^2} \\ \theta &= \tan^{-1}\frac{y}{x}. \end{aligned} \tag{3.40}$$

In polar coordinates the Zernike polynomials are the product of a radial term, which is a function of r, and an azimuthal term, which is a function of θ. In the absence of an ordering scheme, two terms are required to define the order of a particular Zernike polynomial. These terms are the azimuthal order, m, and the radial order, n.

This thesis considers Noll's technique for numbering the Zernike polynomials using a single index rather than the ordered pair, (m,n). This ordering scheme allows the Zernike polynomials to be uniquely identified without the use of of two numbers. These polynomials are defined under Noll's ordering scheme as [38]

$$\begin{aligned} Z_{\text{even } j} &= R_n^m(r)\cos(m\theta), \quad m \neq 0 \\ Z_{\text{odd } j} &= R_n^m(r)\sin(m\theta), \quad m \neq 0 \\ Z_j &= R_n^0(r), \quad m = 0 \end{aligned} \tag{3.41}$$

where the radial polynomials are given by,

$$R_n^m(r) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s(n-s)!}{s![(n+m)/2-s]![(n-m)/2-s]!} r^{n-2s}. \tag{3.42}$$

The azimuthal and radial orders, m and n, are non-negative integers that satisfy $m \leq n$ and $(n - m) =$ even.

Since the Zernike polynomials are orthonormal only on the unit circle, the radius of the aperture must be normalized to apply Zernike polynomials to non-unit radius apertures. The appropriate transformation is [38]

$$\rho = \frac{r}{R}, \tag{3.43}$$

where R is the radius of the aperture. Thus, any wavefront or wavefront phase can be expanded in terms of Zernike polynomials over a circle of radius R as

$$\phi(R_\rho, \theta) = \sum_j a_j Z_j(\rho, \theta) \tag{3.44}$$

where the coefficients a_j are given by [1]

$$a_j = \int \mathcal{W}(\rho)\phi(R_\rho, \theta)Z_j(\rho, \theta)\ \mathrm{d}\rho. \tag{3.45}$$

It is worth noting that because of the underlying Gaussian distributed nature of $W(R_\rho, \theta)$ at every point in space, the random variables a_j are also Gaussian with zero mean and given variance [39].

The first Zernike mode, $Z_1(\rho, \theta)$, is referred to as *piston*, with a coefficient given by

$$a_1 = \int \mathcal{W}(\rho)\phi(R_\rho, \theta)\ \mathrm{d}\rho \tag{3.46}$$

which physically corresponds to the aperture averaged wavefront phase. Single aperture imaging systems are insensitive to piston [30]. Therefore, the piston term is generally omitted in studies of turbulence effects on imaging systems. The piston removed phase, $\varphi(R_\rho, \theta)$, can then be defined [38]

$$\varphi(R_\rho, \theta) = \phi(R_\rho, \theta) - \int \mathcal{W}(\rho)\phi(R_\rho, \theta)\ \mathrm{d}\rho. \tag{3.47}$$

The Zernike modes $Z_2(\rho, \theta)$ and $Z_3(\rho, \theta)$ are referred to as *tip* and *tilt*, respectively. These two modes correspond to the orthogonal components of the wavefront phase in the telescope aperture plane. Other Zernike modes with well known names include $Z_4(\rho, \theta)$, which is referred to as *defocus*; $Z_5(\rho, \theta)$ and $Z_6(\rho, \theta)$ are the orthogonal components of *astigmatism*; $Z_7(\rho, \theta)$ and $Z_8(\rho, \theta)$ are the orthogonal components of *coma*; and $Z_{11}(\rho, \theta)$ is referred to as *spherical aberration*. Figure 3.12 shows the form of the first eight Zernikes polynomials.

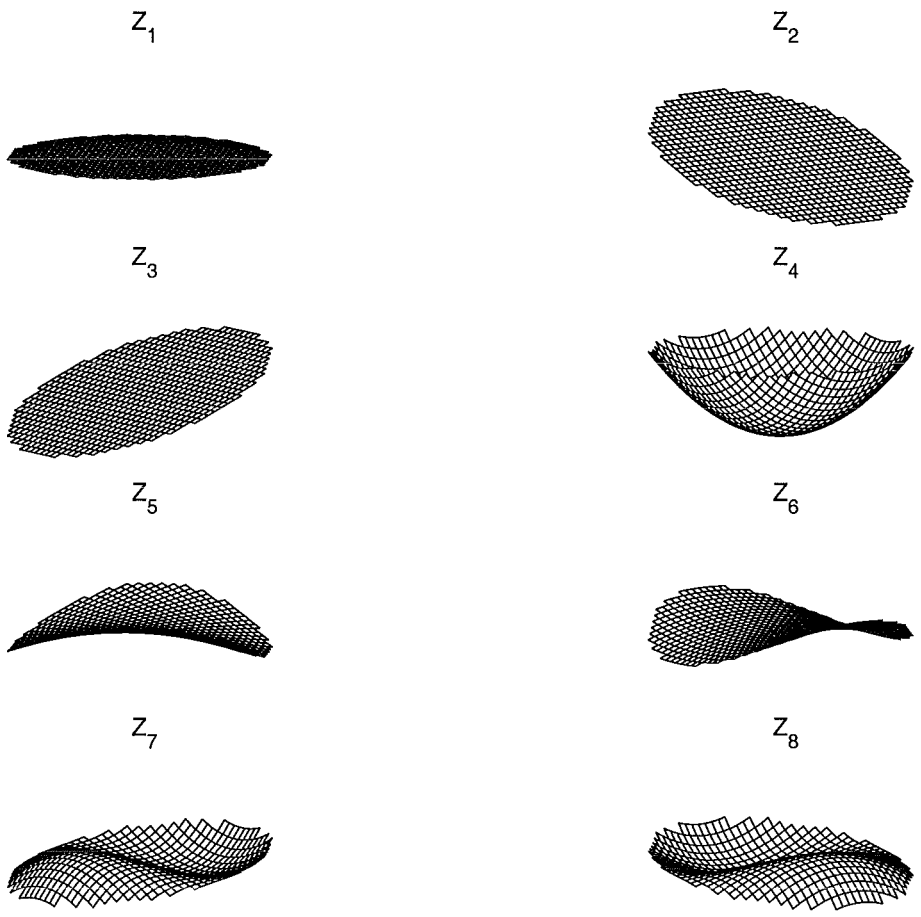


Figure 3.12: *First eight Zernike polynomials.* Z_1 *= piston,* Z_2 *= tip,* Z_3 *= tilt,* Z_4 *= defocus,* Z_5, Z_6 *= astigmatism,* Z_7, Z_8 *= coma.*

Noll [38] also calculated the mean square residual phase error that resulted when the first J Zernike terms are corrected and the results of this calculation are now discussed. The wavefront phase resulting from the summation of the first J Zernike terms is

$$\phi_c = \sum_{j=1}^{J} a_j Z_j(R_\rho, \theta). \tag{3.48}$$

so that the mean squared phase error σ_ϕ^2, or to use Noll's terminology Δ_J, when these J terms are exactly corrected is

$$\begin{aligned} \sigma_\phi^2 &= \Delta_J = \int \mathcal{W}(\rho)\langle[\phi(R_\rho) - \phi_c(R_\rho)]^2\rangle \\ &= \langle\phi^2\rangle - \sum_{j=1}^{J}\langle|a_j|^2\rangle, \end{aligned} \tag{3.49}$$

where $\langle\phi^2\rangle$ is the phase variance, which is infinite for the Kolmogorov spectrum.

In terms of the Strehl ratio, the number of the corrected modes can be computed using (3.21),

$$\begin{aligned} S &= \exp[-\Delta_J] \\ \Delta_J &= -\ln S. \end{aligned} \tag{3.50}$$

For large J the mean square residual phase error can be approximated by [38]

$$\Delta_J \approx 0.2944\ J^{-\sqrt{3}/2} \left(\frac{D}{r_0}\right)^{5/3} \qquad \text{rad}^2 \tag{3.51}$$

and therefore,

$$J \approx \left(\frac{0.2944}{-\ln S}\right)^{\frac{2}{\sqrt{3}}} \left(\frac{D}{r_0}\right)^{5/3}. \tag{3.52}$$

This residual error is an important measure in wavefront estimation, and it is further discussed in the next chapter. Graphs of the required number of Zernike polynomials that have to be corrected as function of D/r_0 to achieve the given Strehl values are plotted in Fig. 3.13. This is done for Strehls of 0.1, 0.4, and 0.8, respectively. Obviously the degree

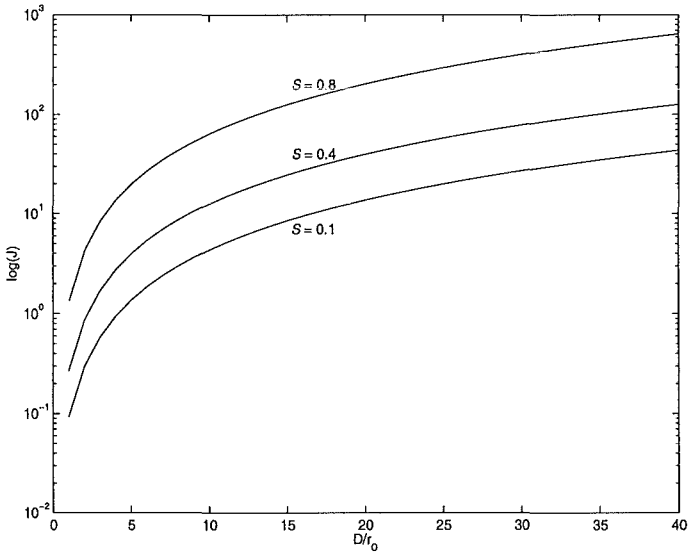


Figure 3.13: *The logarithm (base 10) of the number of Zernike polynomials that must be precisely compensated for, in order to reach Strehl ratios of 0.1 (lower), 0.4 (middle), and 0.8 (upper), as a function of D/r_0.*

of correction required to achieve a certain Strehl value depends very strongly upon D/r_0. For example, if $D/r_0 = 40$, then approximately 1000 Zernike terms must be corrected to

be able to reach the Marechal's criterion (3.1.2). However, if the D/r_0 is reduced to 8, the number of Zernike terms to be corrected decreases significantly to about 100.

As mentioned earlier, the Zernike coefficients are weakly correlated, and therefore the Zernike expansion is not the best expansion to use for Kolmogorov turbulence. Roddier [39], however, shows that a spectral decomposition can be used to express the Karhunen-Loeve expansion in terms of Zernike polynomials, which is now discussed.

The covariance matrix is to be computed from (3.45),

$$\mathbf{C}_a = \langle a_j a_k^* \rangle = \int\int \mathcal{W}(\rho)\mathcal{W}(\rho')Z_j(\rho)C_\phi(R_\rho, R_{\rho'})Z_k^*(\rho)\ \mathrm{d}\rho\mathrm{d}\rho' \tag{3.53}$$

where $C_\phi(R_\rho, R_{\rho'})$ is the phase covariance function

$$C_\phi(R_\rho, R_{\rho'}) = \langle \phi(R_\rho)\phi(R_{\rho'}) \rangle. \tag{3.54}$$

Equation (3.53) can also be written in Fourier space as

$$\langle a_j a_k^* \rangle = \int\int \tilde{Z}_j(k) W\left(\frac{k}{R}\frac{k'}{R}\right) \tilde{Z}_k^*(k')\ \mathrm{d}k\mathrm{d}k', \tag{3.55}$$

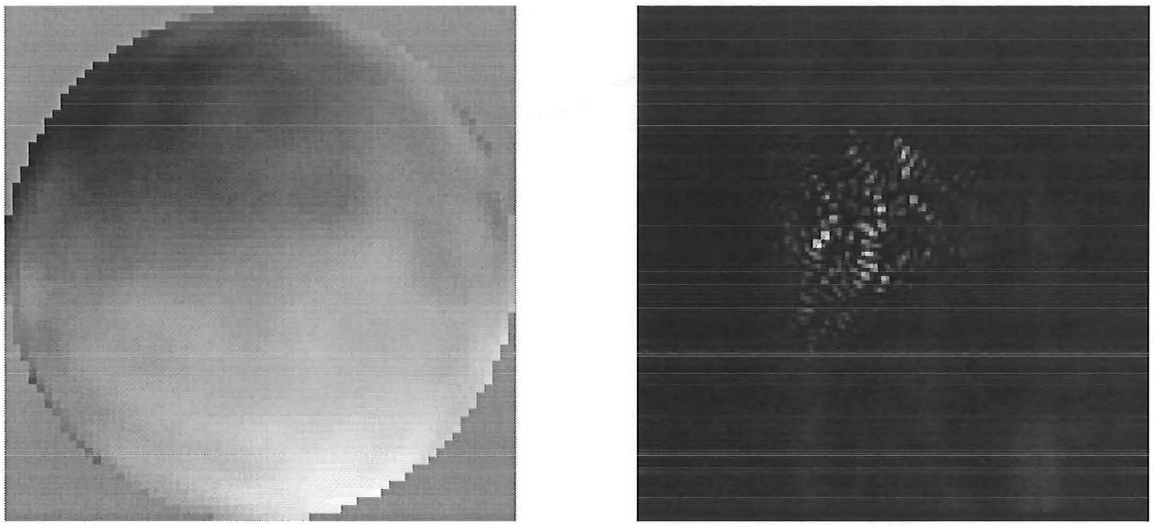
where $\tilde{Z}_j(k) = \mathcal{F}\{Z_j(k)\}$.

Since $\mathbf{C}_a$ is known to be Hermitian, there must exist a unitary matrix $\mathbf{U}$ that diagonalizes it such that $\mathbf{U}\mathbf{C}_a\mathbf{U}^T$ is diagonal. This $\mathbf{U}$ matrix can be computed by using the Singular Value Decomposition (SVD). Gaussian random variables are then generated and therefore they are statistically independent. From here, the desired vector, i.e. the components of the random coefficients for the Zernike polynomials $Z_2, Z_3, \ldots$, can be generated using the relation [39],

$$\mathbf{A} = \mathbf{U}^T\mathbf{B}, \tag{3.56}$$

where $\mathbf{B}$ is a vector containing Gaussian random variables.

Having obtained the $\mathbf{A}$ matrix and using Eq. (3.44), the desired random atmospheric wavefront can then be generated. Figure 3.14a shows a phase screen generated using Zernike polynomials at 300 terms, whilst the corresponding speckle image for $D/r_0 = 10$ is shown in Fig. 3.14b. It is worth noting the difference from the previous approach (Fig. 3.11a) by which it is advantageous to perform a simulation using Zernike polynomials since most applications in optics use a round aperture.



(a) (b)

Figure 3.14: *(a) A typical phase screen produced by using Zernike polynomials (b) The corresponding speckle image ($D/r_0 = 10$).*

3.2.4 Angle-of-arrival statistics

In this section, the concept of angle-of-arrival of incident wavefronts is described. The significance of determining the angle-of-arrival statistics or the mean slope of wavefronts is twofold. First, the correction of the mean slope is an important step in the implementation of a correction system. Second, the Shack-Hartmann wavefront sensing, which is presented in the next chapter, is done by sensing the tilt over sub-apertures.

Recalling Eq. (3.6), the relationship between the wavefront and the phase can be rewritten:

$$W(x,y) = \frac{\lambda}{2\pi}\phi(x,y). \tag{3.57}$$

The angle-of-arrival is then related to the wavefront slope by its first derivative:

$$\begin{aligned} \alpha(x,y) &= \frac{\partial}{\partial x}W(x,y) \\ \beta(x,y) &= \frac{\partial}{\partial y}W(x,y), \end{aligned} \tag{3.58}$$

a point illustrated in Fig. 3.15. Since $\phi(x,y)$ is Gaussian, (α,β) are also Gaussian, which simply means that the image moves around the image plane with a Gaussian probability distribution.

By use of the differentiation property of the Fourier transform (table 2.1) the respective

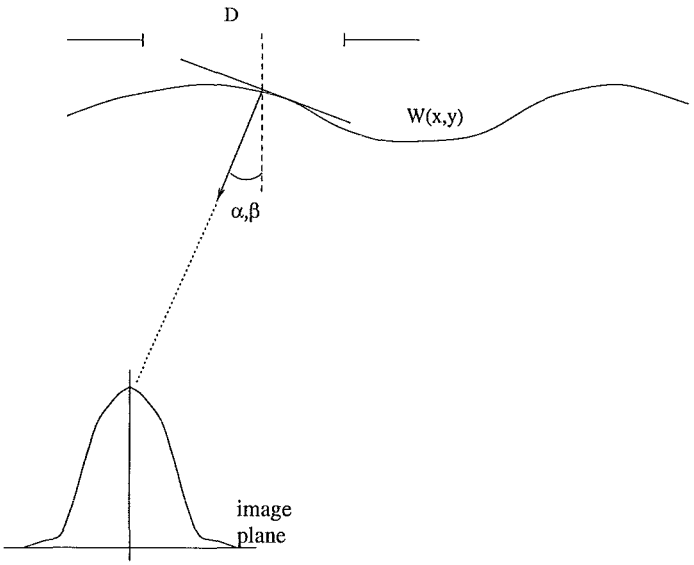


Figure 3.15: *Angle-of-arrival of the incoming wavefront $W(x,y)$ over an aperture D.*

power spectra of the slopes can be expressed as

$$\begin{aligned} \Phi_\alpha(\mathbf{k}) &= \lambda k_x \Phi_\phi(\mathbf{k}) \\ \Phi_\beta(\mathbf{k}) &= \lambda k_y \Phi_\phi(\mathbf{k}), \end{aligned} \tag{3.59}$$

where k_x and k_y are the x and y components of the frequency vector $\mathbf{k}$, $\Phi_\phi(\mathbf{k})$ is the power spectrum of the phase (Eq. 3.38), and Φ_α and Φ_β are the power spectra of the wavefront slopes.

The total variance of the slope is therefore (cf. Eq. 2.44):

$$\sigma^2_{\alpha,\beta} = \lambda^2 \int_D |\mathbf{k}|^2 \Phi_\phi(\mathbf{k})\ \mathrm{d}\mathbf{k}. \tag{3.60}$$

A rigorous analysis of this problem was carried out by Fried [41] to obtain the variance of wavefronts' slopes

$$\begin{aligned} \sigma^2_{\alpha,\beta} &= 0.36 r_0^{-5/3} \lambda^2 D^{-1/3} \\ &= 0.36 \left(\frac{\lambda}{D}\right)^{1/3} \left(\frac{\lambda}{r_0}\right)^{5/3} \\ &= 0.36 \left(\frac{\lambda}{D}\right)^{2} \left(\frac{D}{r_0}\right)^{5/3} \qquad [\mathrm{rad}^2]. \end{aligned} \tag{3.61}$$

It is important to note here that as $D \to 0$, $\sigma^2_{\alpha,\beta}$ goes to infinity for ideal Kolmogorov turbulence. In reality there is an inner scale in the order of mm that limits the variance.

Furthermore, there is no dependence of $\sigma^2_{\alpha,\beta}$ on the wavelength since r_0 is proportional to $\lambda^{6/5}$. The variance given in Eq. (3.61) is of importance since it also describes the variance of the image motion on a detector, a point to be analyzed in chapter 4.

3.2.5 Long-exposure imaging

Long exposure imaging is an imaging experiment in which the captured images are the result of a large number of pupil plane perturbation random processes (section 2.3.5). As discussed in section 3.1, the wave motion is characterized by a complex-valued scalar function of position and time. This implies that atmospheric turbulence effects evolve with time. This temporal evolution is characterized by a correlation time τ. This can be thought of as the atmosphere generating a new field perturbation realization every τ seconds. If an imaging system exposes the image plane detector for a large number of atmospheric correlation times, the image is called a long exposure image [33]. In long exposure images it is necessary to compensate for the object's apparent motion by moving the telescope appropriately. The image obtained by tracking the object is thus a time average

$$\langle d(x,y)\rangle = \langle f(x,y)\rangle \odot \langle h(x,y)\rangle + \langle n(x,y)\rangle. \tag{3.62}$$

The major difference between $h(x,y)$ and $\langle h(x,y)\rangle$ is best discussed in terms of their Fourier transforms. In $\mathcal{H}(u,v)$ the spatial frequencies above a certain frequency and their phases are effectively randomized. When a time average is performed these spatial frequencies tend to cancel. Loss of these higher spatial frequencies causes a point source to be smeared [10].

In order to compute the long exposure transfer function it is convenient to rewrite the pupil function or the complex amplitude P in terms of its magnitude and phase in polar form as

$$P = P_{opt}U(x,y) \tag{3.63}$$

where $U(x,y)$ is given in (3.32), and the optical pupil function for a perfect telescope is defined by,

$$P_{opt} = \begin{cases} 1 & \text{inside the aperture} \\ 0 & \text{outside the aperture.} \end{cases}$$

Recalling the transfer function for incoherent imaging in Eq. (3.28), the long exposure transfer function can be derived as,

$$\langle \mathcal{H}(u,v)\rangle_{LE} = \langle \int\int P^*_{opt}(\lambda z u', \lambda z v')U^*(\lambda z u', \lambda z v')$$

$$\begin{aligned} P_{opt} \quad & (\lambda z u' + \lambda z u, \lambda z v' + \lambda z v) U(\lambda z u' + \lambda z u, \lambda z v' + \lambda z v) \mathrm{d}u' \mathrm{d}v' \rangle \\ = \quad & C_U(\lambda z u, \lambda z v) T_{opt}(u, v) \end{aligned} \tag{3.64}$$

where $T_{opt}(u, v)$ is the telescope transfer function

$$T_{opt} = \int\int P^*_{opt}(\lambda z u', \lambda z v') P_{opt}(\lambda z u' + \lambda z u, \lambda z v' + \lambda z v) \mathrm{d}u' \mathrm{d}v', \tag{3.65}$$

and C_U is the atmospheric transfer function which can be computed by use of Eq. (3.33):

$$C_U(\mathbf{u}) = \exp[-\frac{1}{2}\mathcal{D}_\phi(\mathbf{u})]. \tag{3.66}$$

The long exposure transfer function is therefore

$$\langle \mathcal{H}(u, v) \rangle_{LE} = T_{opt}(u, v) \exp\left[-3.44 \left(\frac{\lambda z |\mathbf{u}|}{r_0}\right)^{5/3}\right], \tag{3.67}$$

or assuming a circular telescope aperture [41]

$$\langle \mathcal{H}(u, v) \rangle_{LE} = \frac{2}{\pi}\left[\cos^{-1}(\mathbf{u}) - \mathbf{u}\sqrt{1 - \mathbf{u}^2}\right] \exp[-3.44(D/r_0)^{5/3}\mathbf{u}^{5/3}]. \tag{3.68}$$

3.2.6 Short exposure transfer function

Short exposure imaging refers to the situation in which the exposure time is short enough to freeze the effect of the atmosphere. In this case the wavefront tilt has no effect on image quality [33], but does cause the image to be displaced from the centroid. The short exposure image is distorted because the wavefront tilt component is accompanied by higher spatial frequency perturbations. Fried [41] has shown that tilt accounts for most of the power in the phase perturbations. Therefore, good performance of an imaging system can be achieved if the random shift of the image is compensated for.

To compensate for the random shift of the image, the imaging system takes short exposure images. Each short exposure image is subsequently re-centroided, thereby effectively removing the image shift caused by the random tilt. The ensemble of images is then summed. The summing of the centroided images is equivalent to the long exposure imaging described above except no centroiding is performed for the long exposure case. The transfer function for this type of imaging is referred to as the short exposure OTF. It is important to remember that a reference to a short exposure OTF is a reference to a long exposure OTF for the case of tilt compensation.

The short exposure OTF was first derived by Fried [41] and detailed explanations can be found in the book by Roggemann and Welsh [1]. The approximate analysis for the short exposure OTF assuming a circular aperture is [41]

$$\langle \mathcal{H}(u,v) \rangle_{SE} = T_{opt}(u,v) \exp\left[-3.44 \left(\frac{\lambda z |\mathbf{u}|}{r_0} \right)^{5/3} \left[1 - \alpha \left(\frac{\lambda z |\mathbf{u}|}{D} \right)^{1/3} \right] \right]. \tag{3.69}$$

3.2.7 Long versus short exposure

The OTFs and PSFs encountered for long and short exposure images are remarkably different as illustrated in Fig. 3.16. Figure 3.16a shows an ideal OTF for the normalized aperture dimension, and the corresponding diffraction-limited PSF is shown in Fig. 3.16b. The OTF of a short exposure image has significant fluctuations of both magnitude and phase as a function of spatial frequency, and the corresponding PSF is no longer an Airy function. Both are shown in Figs. 3.16c and 3.16d, respectively. Recall that because of the averaging process in the long exposure case, the higher frequencies are cancelled out, and therefore produces a narrow function as depicted in Fig. 3.16e. Taking the inverse Fourier transform of this function gives a smooth and broad PSF (Fig. 3.16f).

The difference between the long and short exposure results lie in the effect of the term $\left[1 - \alpha \left(\frac{\lambda z |\mathbf{u}|}{D} \right)^{1/3} \right]$. In the long exposure case $\alpha = 0$, and so this term reduces to unity. In the short-exposure case, nonzero values for α, which is different in the near-field and far-field cases (cf. Eqs. (3.10) and (3.11)), show that the tilt component of phase has no effect on the OTF, and phase plays a less important role in the far-field case than the near-field case. In the near-field case, all of the blur comes from phase effects, whereas in the far-field case, only half of the blur arises from phase perturbations, with the other half arising from amplitude effects [33].

One of the most important facts about short exposure images is that their quality is unaffected by the tilt component of the wavefront distortions. A tilt of the incoming wavefront only shifts the center of the image and does not affect the image quality. In contrast, for long exposure images, changing the tilt of the incoming wavefront serves to broaden the PSF and narrow the OTF.

In order to summarize the effects of the turbulence, Fried [41] defined the normalized reso-

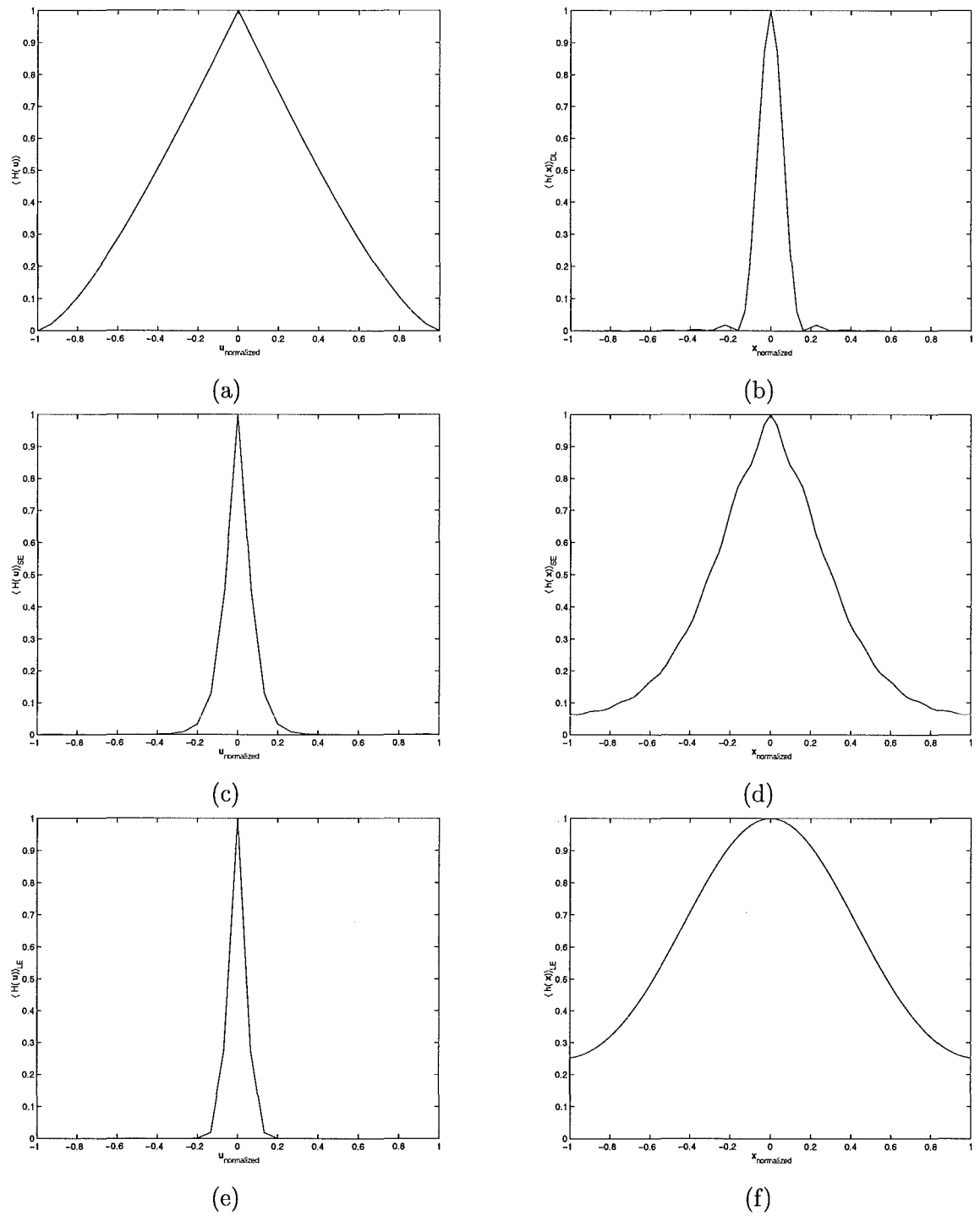


Figure 3.16: *Typical OTFs and PSFs for (a),(b) ideal case; (c),(d) short exposure imaging; (e),(f) long exposure imaging. The ratio D/r_0 is set equal to 8, and the aperture dimensions are normalized to 1.*

lution of the short and long exposure transfer functions as

$$\frac{\mathcal{R}}{\mathcal{R}_{max}} \equiv \int_{-\infty}^{\infty} \langle \mathcal{H}(\mathbf{u}) \rangle \, \mathrm{d}\mathbf{u} \tag{3.70}$$

obtaining the results shown in Fig. 3.17. $\mathcal{R}_{max}$ is the limiting value of the long exposure

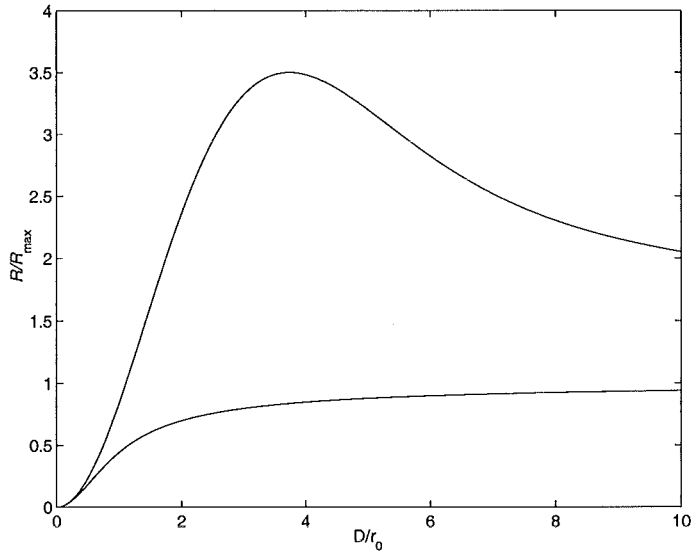


Figure 3.17: *The resolution $\mathcal{R}$ normalized to the value $\mathcal{R}_{max}$ at $D/r_0 \to \infty$, for long-exposure (lower curve) and short-exposure (upper curve) imaging. Note the peak value of $\mathcal{R}/\mathcal{R}_{max} \approx 4$ at $D/r_0 \approx 4$.*

resolution as the aperture becomes arbitrarily large. The value of $\mathcal{R}_{max}$ is [41]

$$\mathcal{R}_{max} = (\pi/4)(r_0/\lambda f_L)^2 \quad \text{cycles}^2/\text{m}^2 \tag{3.71}$$

where f_L is the focal length of the aperture.

As can be seen in Fig. 3.17, the long exposure resolution approaches an asymptotic limit $\mathcal{R}_{max}$ which cannot be exceeded no matter how large the aperture dimension. This implies that atmospheric turbulence places an absolute upper limit on the resolution that can be obtained with long exposure through the atmosphere. It is also obvious in Fig. 3.17 that significantly better resolution may be obtainable with a short exposure than with a long exposure. A maximum of ≈ 4 for the short-exposure case occurs at $D/r_0 \approx 4$. Thus, for a simple tip-tilt system, the optimum value of D/r_0 is approximately 3.8, and an improvement in resolution of about four is obtained. However, the improvement in resolution depends strongly on how resolution is defined, and the value ≈ 4 above is only for the particular definition of a resolution.

3.3 Laser Guide Star

With the deployment of adaptive optics systems (section 1.2.2) on various astronomical telescopes, it is hoped that in the near future astronomers will be able to regularly obtain images for which the degrading effects of atmospheric turbulence have been, at least, partially corrected. A problem remains, however, that such systems require a bright reference source close to the object being observed to allow accurate measurement of the instantaneous distorted wavefront. Because of this restriction, for visible wavelengths in particular, this means that only a very small fraction of the sky can be covered using natural objects, and therefore only a relatively small number of interesting astronomical targets can be imaged using adaptive optics systems within this wave-band.

A possible solution to the problems associated with natural guide stars is the creation of an artificial beacon using a laser to project a spot in the atmosphere above the telescope. Artificial guide stars allow access to nearly the entire sky even at visible wavelengths since the laser can be pointed in any direction. Within the region where the beam is focused some of the light is scattered from the atmospheric constituents, and some of this scattered light is returned in the direction of the telescope. The backscattered light is used by the wavefront sensor to measure the turbulence-induced aberrations. Ideally the use of laser guide stars eliminates the fundamental problem of low light levels in the wavefront sensor since the brightness of the beacon is controlled by the laser power.

Two main approaches for the generation of laser beacons involve the use of Rayleigh scattering in the stratosphere and the use of resonance fluorescence of sodium (Na) atoms in the mesosphere [52]. Beacons employing Rayleigh scattering are called Rayleigh beacons, while guide stars employing Na resonance fluorescence are called Na beacons. Rayleigh guide stars are most efficiently created at relatively low altitudes, ranging from 10 to 20 km. On the other hand, Na beacons are located at an altitude of 90 km [52]. Because the sodium excitation is resonant for Na beacons, a laser can create a beacon more than 10,000 times brighter than a Rayleigh system of equal laser power [13]. This is one reason the sodium beacon is preferable to the Rayleigh beacon.

3.3.1 Cone effect and angular anisoplanatism

The advantage of employing a laser beacon is that it can be positioned potentially anywhere in the sky and therefore any object can be studied; the disadvantages include the large

expense involved (due to the high power laser required) and the fact that the beacon must be created within the atmosphere itself, therefore higher layers of turbulence are less well sampled. This problem is known as the *cone effect* in adaptive optics. It means that the optical path through the atmosphere from the astronomical source and from the artificial source are significantly different [50]; wavefronts restored from the artificial source image must be corrected for that effect before being used to compensate the object wavefront. Figure 3.18 illustrates the cone effect. The volume V_l made by the artificial star, which is a

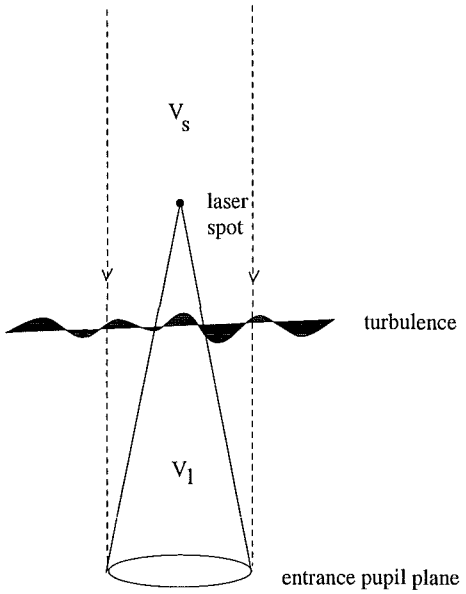


Figure 3.18: *The cone effect.*

volume of a cone, is clearly different from the volume of a cylinder produced by astronomical objects.

Another significant problem in laser guide star techniques is called the angular anisoplanatism [50]. The guide star is, in general, angularly displaced from the science object, and therefore, the wavefront sampled by the guide star is not the same as that from the science object because of this angular displacement, a point illustrated in Fig. 3.19. When the wavefronts are the same, then the system is isoplanatic, and when they are not the same it is referred to as *anisoplanatic*. There is an angle θ_0, called the isoplanatic angle, which quantifies the region over which there is approximate isoplanatism, it is given by [50]

$$\theta_0 = \left[2.91\left(\frac{2\pi}{\lambda}\right)^2 (\cos\gamma)^{-1}\int_0^\infty C_N^2(z)z^{5/3}\mathrm{d}z\right]^{-3/5} \quad \text{rad} \tag{3.72}$$

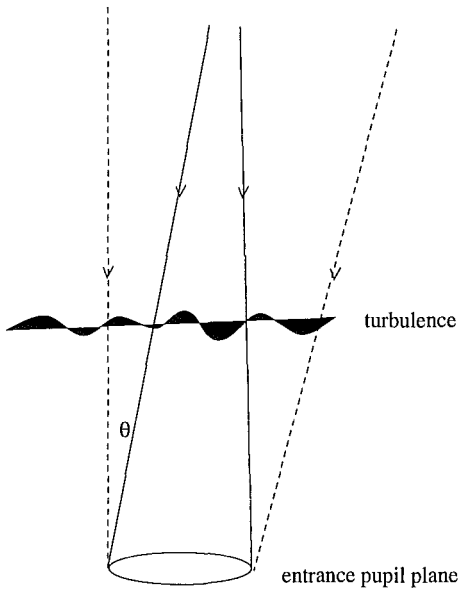


Figure 3.19: *Angular isoplanatism which is one of the greatest restriction to the effectiveness of adaptive optics systems.*

and the mean squared wavefront error σ_θ^2 for a science object observed θ away from the guide star is

$$\sigma_\theta^2 = \left(\frac{\theta}{\theta_0}\right)^{5/3} \quad \text{rad}^2. \tag{3.73}$$

Note that like the Fried parameter, r_0, the isoplanatic angle is proportional to $\lambda^{6/5}$ (cf. Eq. (3.36)).

3.3.2 Phase retrieval by means of laser guide star

This sub-section introduces an analysis of the performance of an adaptive optics system using laser guide stars, in terms of the corrected wavefront phase. Providing a high speed computer, phase retrieval seems very promising. The detail how this can be the case is described in chapter 5.

Telescopes equipped with adaptive optics system require the use of laser spots, backscattered by a high altitude atmospheric layer, as bright reference sources at visible wavelengths. A sufficiently bright reference source is necessary to achieve the correction of the wavefront phase distortion.

However, there are some serious drawbacks with this technique. As well as the problems

associated with the cone effect and angular anisoplanatism discussed earlier, a more severe problem is the tilt determination problem [52]. The overall tip-tilt of the wavefront is not measurable with a laser guide star, as a consequence of the round trip of the light (upward and downward propagation). This insensitivity to wavefront tilt for an artificial beacon implies that the adaptive optics system must employ a natural star to determine the wavefront tilt.

As noted in section 3.2 the tilt is completely determined by the 2^{nd} and 3^{rd} modes in the Zernike polynomials, and that the tilt does not affect the image quality. This implies that a technique like phase retrieval can still be applied to compensate for higher order distortions. Furthermore, the usual approach in an adaptive optics system is to feed the light coming from a laser guide star into a wavefront sensor which analyses the perturbations induced by the turbulent atmosphere. Both the finite spatial sampling of the wavefront sensor and the finite degrees of freedom of the deformable mirror result in the same general type of system performance degradation.

As presented in chapter 5, phase retrieval is a potential alternative to wavefront sensors which are discussed in the next chapter. It is shown that modes higher than tilt are perfectly corrected when using prior information in phase retrieval. This determines the ultimate quality of the corrected image.

Chapter 4

WAVEFRONT SENSING

The principal effect of the turbulence is a continuous distortion of the phase of the wavefront in the aperture of a telescope. This phase distortion significantly degrades the resolution of images formed by the telescope. It can be corrected using adaptive optics and as a consequence, the problem of estimating the wavefront phase is fundamental to accurate astronomical image reconstruction. This chapter considers the fundamental limits of the wavefront sensors that are critical component in adaptive optics systems.

Two main techniques are considered to estimate the degradation of the wavefront phase; wavefront sensing [42–45] and direct phase retrieval [67–69]. The former is based on the measurement of the slope, or the curvature of the phase function on the receiving pupil, simplifying the problem in a manner that results in a linear problem. These techniques all use quantities averaged over a region and thus have the drawback that there is some destruction of the available information. As a result, they all produce a sub-optimal estimate of the wavefront. The latter, phase retrieval, is complicated by the non-linear relationship between phase aberrations and the measurements. However, because it utilizes all the available information, phase retrieval has a greater potential performance, and is discussed in detail in the next chapter.

Since a direct measurement of the wavefront phase is not possible at optical wavelengths, devices such as wavefront sensors indirectly measure an average of the phase over regions. In particular, these sensors are generally sensitive to the derivatives of the wavefront phase.

The wavefront phase is then obtained by using the measured wavefront derivatives in a phase reconstruction algorithm.

A wavefront sensor is connected to the deformable mirror (see section 1.2.2) through the actuator command computer. The purpose of the computer is to take measurements from the wavefront sensor and map them into real time control signals for the deformable mirror. The wavefront reconstruction is then done by controlling the surface of the deformable mirror. The goal of this system is to achieve the ideal performance limit: diffraction-limited imaging. In this case the imaging system would operate as if in an aberration free environment and only the diffraction effects discussed in the previous chapter limit the performance.

Unfortunately theoretical and experimental evidence [42,43,45] have shown that true diffraction-limited performance is impossible to achieve. The limitations of this process are twofold. First, there is the ability of the wavefront sensor to measure the wavefront. Second, there is the ability of the mirror to replicate the true wavefront. The former is fundamentally limited by the available light for measurement, whilst the latter is a technological limitation. Since the aim of this study is to investigate the fundamental limitations of wavefront sensing, the actuator mirror is assumed to be perfect in its ability to produce a desired shape. Its performance can thus be modeled as a sum of an infinite series of orthogonal basis functions. In addition, the wavefront sensing and compensation by the mirror are assumed to happen instantaneously.

Before commencing the detailed discussions of a particular wavefront sensor, the three types of wavefront sensor in the field of astronomy are first presented in general in section 4.1. Section 4.2 provides more detail on the theory of the Shack-Hartmann sensor which is the most commonly used. This background includes discussions of the wavefront estimation and also introduces the practical limitations of the sensor. A conventional analysis of this sensor is presented in section 4.3, whilst section 4.4 introduces a new analysis of the measurement errors in the Shack-Hartmann sensor. The latter differs from previously published work in the estimation of the noise inherent in the centroid calculation used in this sensor. Simulations are described in section 4.5 to demonstrate the validity of the derived theory.

4.1 Wavefront Sensor Types and Models

Wavefront sensors used in adaptive optics systems do not directly measure the wavefront phase $\phi(u, v)$. Instead, the wavefront sensors measure the spatial gradient or Laplacian of $\phi(u, v)$. The spatial gradient is commonly referred to as the wavefront slope while the Laplacian is usually referred to as the wavefront curvature. An estimate of $\phi(u, v)$ is computed from either the wavefront slope information or the wavefront curvature by use of a phase reconstruction algorithm. The three main types of wavefront sensors used in adaptive optics systems are now reviewed. All of these wavefront sensors are sensitive to the derivatives of the wavefront phase and have been used in existing adaptive optics system [1, 4, 18].

4.1.1 Shack-Hartmann wavefront sensor

The Shack-Hartmann wavefront sensor (SHWS) is the most commonly used sensor in astronomical imaging to date. A diagram showing the relationship of telescope aperture with the SHWS is depicted in Fig. 4.1a. The telescope aperture D is firstly reimaged at a smaller

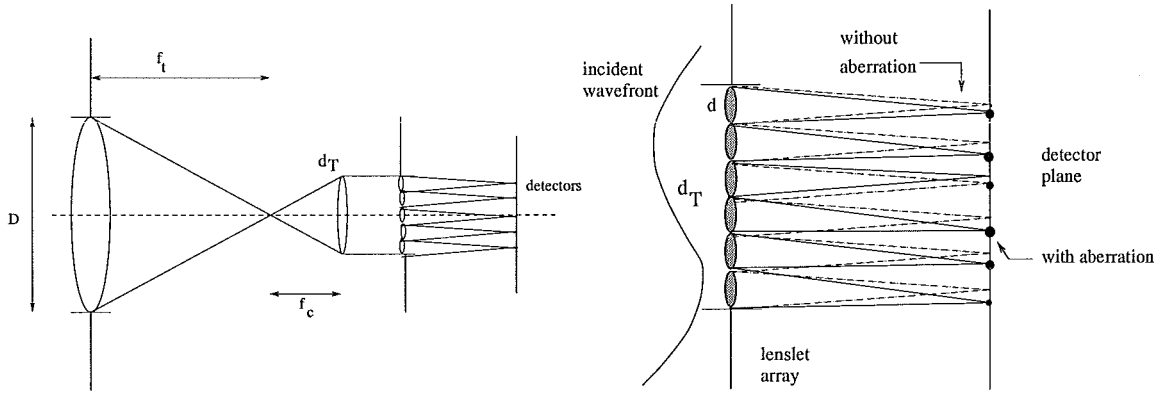


Figure 4.1: *(a) Relation between a telescope aperture and lenslets (b) Diagram of the Shack-Hartmann WFS*

scale before being subdivided by a lenslet array. The relationship between the main and second apertures can be expressed mathematically:

$$D = \frac{f_t}{f_c} d_T \tag{4.1}$$

where d_T is the size of the smaller aperture as a total size of a collection of the lenslets area, and f_t and f_c are the focal lengths of the aperture D and d_T, respectively.

The SHWS spatially segments the incident wavefront with a lenslet array of size d [42,43]. Each segment is referred to as a sub-aperture and focuses a spot onto a detector in the focal plane, a point illustrated in Fig. 4.1b. The spot is shifted by a quantity proportional to the local slope of the wavefront, a point described in more detail in section 4.2. The array of these shifted spots is recorded by a CCD at the same time as the image is recorded. The shifts are calculated by a centroiding algorithm in two perpendicular directions. The relationship of these shifts to the mean slope of the phase aberration over the corresponding lenslets is detailed in section 4.2.1.

4.1.2 Shearing interferometer

The shearing interferometer is also based on measuring the wavefront slope. In this technique, the wavefront is combined with a shifted version of itself to form interference fringes [4]. As shown in Fig. 4.2, the two images formed interfere in their overlapping area. Placing

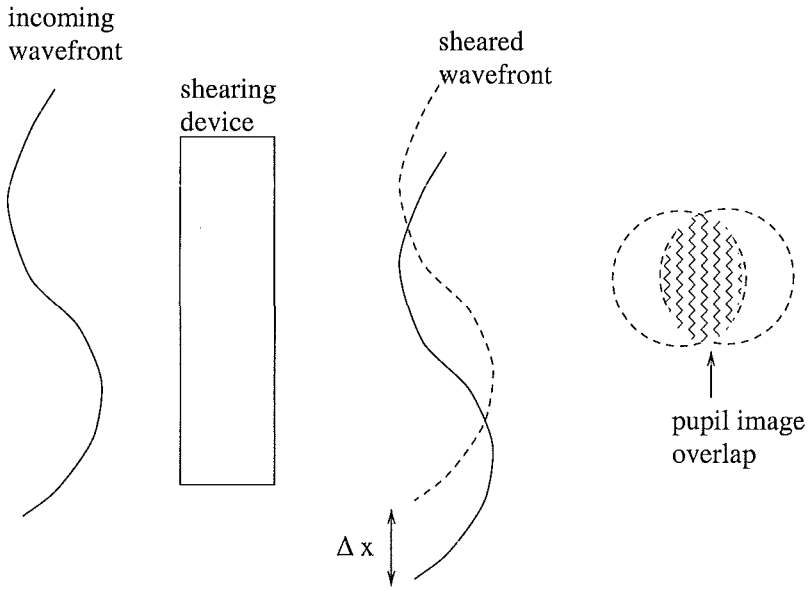


Figure 4.2: *Diagram of the shearing interferometer. The adjustable parameter* $\Delta\mathbf{x}$ *corresponds to the pupil image overlap denoted by the shaded area.*

the detector in a pupil image plane, the observed intensity is given by

$$\begin{aligned}[I(x,y)]_{\Delta\mathbf{x}} &= |\exp[ikW(x,y)] + \exp[ikW(x+\Delta x, y+\Delta y)]|^2 \\ &= 2 + 2\cos[k(W(x,y) - W(x+\Delta x, y+\Delta y)]\end{aligned}$$

$$= 2 + 2\cos\left[k\frac{\partial W}{\partial \mathbf{x}}\right], \tag{4.2}$$

where $\frac{\partial W}{\partial \mathbf{x}}$ is the wavefront slope in two orthogonal directions.

In order to determine completely the wavefront, it is necessary to produce two interferograms sheared in orthogonal directions x and y. The sheared beams are superimposed on an optical detector to form an interference pattern which can then be used to estimate the average wavefront slope. Shearing interferometry is a technologically complicated technique, but it is flexible in choice of $\Delta\mathbf{x}$. This adjustable parameter for the amount of shear can be used to optimize the measurement error [4, 45]. The optimum shear is depending on the object and has been approximated by Welsh *et al* [45] as $\Delta\mathbf{x}_{\text{opt}} = 1/(\sqrt{2}\pi\beta)$, where β is the angular size of the object in radians. The larger the object size β, the smaller $\Delta\mathbf{x}$, and the larger the fringe visibility. This optimization of shear with the observing conditions is very important in order to obtain the best performance of this interferometer. For further details, the reader is referred to [1, 4, 45].

4.1.3 Curvature sensor

The other type of wavefront sensor that has found use in adaptive optics systems is the curvature sensor [49]. This sensor measures the wavefront curvature instead of wavefront slope measurement. This wavefront curvature is computed by the Laplacian of the wavefront surface $\phi(u, v)$ by

$$\nabla^2\phi = \frac{\partial^2\phi}{\partial u^2} + \frac{\partial^2\phi}{\partial v^2} \tag{4.3}$$

which is also well known as the Poisson equation [49]. The principle of this sensor is illustrated in Fig. 4.3. A telescope with pupil D and of focal length f_L images the source in its focal plane. The curvature sensor consists of two image detectors placed out of focus. The first detector records the irradiance distribution in plane P_1 at a distance f before the focal plane. The second one records the irradiance distribution in plane P_2 at the same distance behind the focus. It is important to note that the images at P_1 and P_2 are sufficiently displaced from the image plane to be like blurred views of the aperture, not defocused versions of the image. A local wavefront curvature produces an excess of illumination in one plane and a lack of illumination in the other. The difference between the two irradiance distributions is a measure of the local wavefront curvature.

Roddier [49] shows that the normalized difference between the illuminations in planes P_1

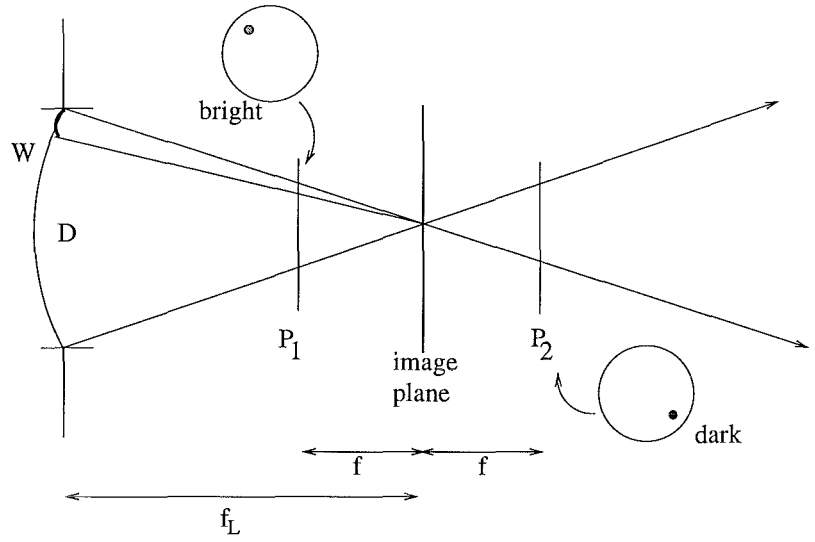


Figure 4.3: *Diagram of the curvature sensor. The difference between the illumination at P_1 and P_2 is a measure of the local wavefront curvature.*

and P_2 is given by

$$\frac{I_2(\mathbf{x}) - I_1(\mathbf{x})}{I_2(\mathbf{x}) + I_1(\mathbf{x})} = \frac{f_L(f_L - f)}{2l}\left[\frac{\partial\phi(\mathbf{u})}{\partial\mathbf{u}}\delta_c + \nabla^2\phi(\mathbf{u})\right] \tag{4.4}$$

where δ_c is a linear impulse distribution around the pupil edge weighted by the wavefront radial tilt $\partial\phi/\partial\mathbf{u}$. This term accounts for any overall tilts in the wavefront which have the effect of shifting the observed intensity. The sensor provides all the information needed to reconstruct the wavefront, using a set of linear equations that can be derived in a similar manner to the Shack-Hartmann sensor, a process discussed further in the next section.

One of the advantages of using a curvature sensor is that solving the Poisson equation can be accomplished directly by applying the Laplacian measurements to a membrane or bimorph mirror [49, 60]. Since curvature sensors are still not in widespread use in adaptive optics systems, further analysis of this sensor has not been carried out. The reader is referred to [49] for further details.

4.2 Shack-Hartmann Measurement Model

Of the previously presented wavefront sensors, to date the most common type sensor in use is the Shack-Hartmann sensor because of its simplicity, flexibility, and widespread application. Whether it is in fact optimal is a point in contention. The scope of the rest of this chapter is,

therefore, limited to an analysis of the performance of this type of sensor in its most common configuration. The original work in this field was performed by a number of authors, including Wallner [9], Primot *et al* [42], Welsh and Gardner [43], Fried [56], Southwell [57], and Hermann [58].

4.2.1 Tilt sensing

Before commencing the Shack-Hartmann discussion, the theory of tilt sensing is reviewed since this is fundamental to how the sensor works. The wavefront tilt induced on the object field is caused by the turbulence. Since such a large portion of the wavefront aberrations is associated with wavefront tilt as discussed in section 1.2.2, provided the aperture is small, it is a good approximation to the phase aberration across a small aperture.

Consider an aperture with a diameter D shown for one-dimensional wavefront, which is similar to the two-dimensional case, in Fig. 4.4. The centroid position of the instantaneous

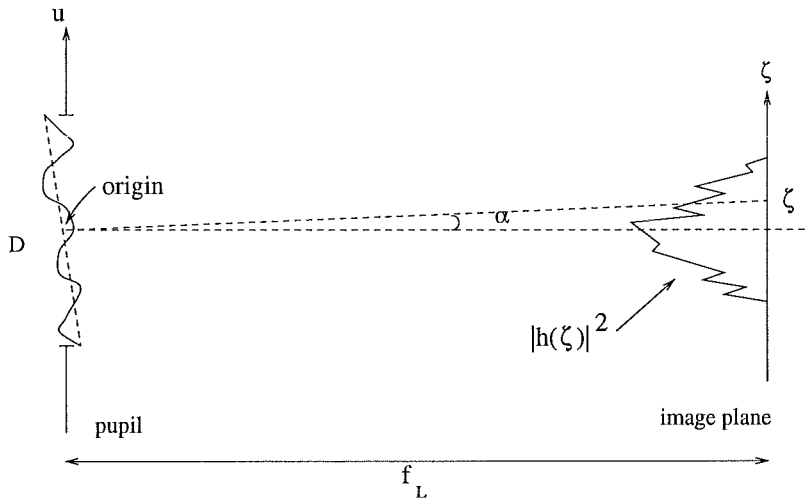


Figure 4.4: *Tilted wavefront with its instantaneous image. The centroid of the image is mainly determined by the slope of the incident wavefront (α).*

image with profile $|h(\zeta)|^2$ is defined as [42]

$$\begin{aligned} \zeta &= \frac{\int |h(\zeta)|^2 \, \zeta \, \mathrm{d}\zeta}{\int |h(\zeta)|^2 \, \mathrm{d}\zeta} \\ &= \lambda f_L \frac{\int |h(\zeta')|^2 \, \zeta' \, \mathrm{d}\zeta'}{\int |h(\zeta')|^2 \, \mathrm{d}\zeta'}, \end{aligned} \tag{4.5}$$

where $\zeta' = \frac{\zeta}{\lambda f_L}$ and the complex image $h(\zeta)$ is the Fourier transform of the pupil function as given by Eq. (3.11). Assuming the magnitude of the pupil function to be unity, by using Parseval's theorem (table 2.1) it is straightforward to derive

$$\begin{aligned} \int |h(\zeta')|^2 \ \mathrm{d}\zeta' &= \int |H(u)|^2 \ \mathrm{d}u \\ &= D. \end{aligned} \tag{4.6}$$

In addition, from the geometry shown in Fig. 4.4, the slope of the incident wavefront can be found as $\alpha = \frac{\zeta}{f_L}$. Using Eqs. (4.5) and (4.6), and the properties of Fourier transform described in section 2.2, the slope of the wavefront can then be derived

$$\begin{aligned} \alpha &= \frac{\lambda}{D} \int |h(\zeta')|^2 \zeta' \ \mathrm{d}\zeta' \\ &= \frac{\lambda}{D} \left(-\frac{1}{2\pi i}\right) \frac{\partial}{\partial u} \{H * H^*\} \mid_{origin} \\ &= \frac{\lambda}{D} \left(-\frac{1}{2\pi i}\right) \frac{\partial H}{\partial u} * H^* \mid_{origin}. \end{aligned} \tag{4.7}$$

The notation $\mid_{origin}$ is used to denote that only the origin of the aperture is taken since the slope is computed with respect to this point (see Fig. 4.4). Recall that the OTF is determined by the autocorrelation (section 2.2). Assuming that the one of the dominant effects of the atmosphere is random wavefront phase [55], and the magnitude is uniform in the pupil, the OTF can be derived as

$$H * H^* = \int_{-\frac{D}{2}+\frac{\Delta u}{2}}^{\frac{D}{2}-\frac{\Delta u}{2}} \exp[ikW(u - \frac{\Delta u}{2})]\exp[-ikW(u + \frac{\Delta u}{2})] \ \mathrm{d}u. \tag{4.8}$$

The derivative of Eq. (4.8) with respect to u is

$$\left\{\frac{\partial H}{\partial u} * H^*\right\} = ik \int_{-\frac{D}{2}}^{\frac{D}{2}} \exp[ikW(u - \frac{\Delta u}{2})] \ \frac{\partial W(u - \frac{\Delta u}{2})}{\partial u} \exp[-ikW(u + \frac{\Delta u}{2})] \ \mathrm{d}u. \tag{4.9}$$

Since only the origin of the aperture is needed in Eq. (4.7), Δu can be set to 0, and thus Eq. (4.9) results in

$$\left\{\frac{\partial H}{\partial u} * H\right\}_{\Delta u=0} = ik\left[W\left(\frac{D}{2}\right) - W\left(-\frac{D}{2}\right)\right]. \tag{4.10}$$

Substituting Eq. (4.10) into Eq. (4.7) gives the mean slope estimate

$$\alpha = \frac{W\left(-\frac{D}{2}\right) - W\left(\frac{D}{2}\right)}{D}. \tag{4.11}$$

Hence, the angle determined by the centroid (or the first moment) is simply related to the mean slope of the phase across the pupil, a point illustrated in Fig. 4.5.

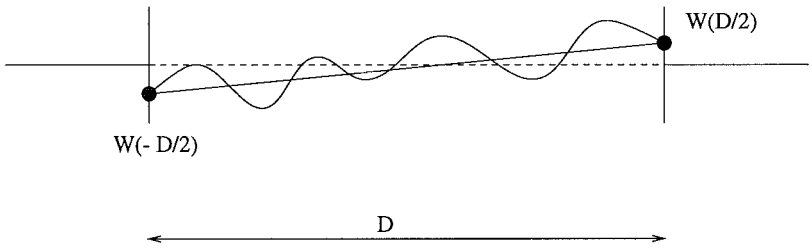


Figure 4.5: *A one-dimensional representation of wavefront tilt. The straight line between two dots represents the mean slope estimate.*

It should be noted, however, owing to the approximation given in Eq. (4.11), it may happen that if $W(-D/2) - W(D/2) = 0$, then the wavefronts produce no centroid motion, a situation depicted in Fig. 4.6. Figure 4.6 shows two incident wavefronts with different mean slopes over the aperture D. The estimated mean slopes are shown in dashed lines, each having a "sub-aperture" of width x and y, respectively. Let z be the distance between the aperture and the intersection of the two wavefronts. Two sub-images, each corresponding to one wavefront, are thereby formed behind the aperture at a focal length f_L. For a perfectly flat, tilt free plane wave, all sub-images fall directly on the optical axes of the aperture as denoted in the figure by ζ_0. For aberrated wavefronts, each sub-image is displaced from ζ_0 by an amount $f_L \times \alpha$, where α is the mean slope of the wavefront. Assuming uniform magnitude across the aperture, the displacements can be approximated:

$$\begin{aligned} \zeta_1 &= -f_L \times (z/y) \\ \zeta_2 &= f_L \times (z/x), \end{aligned} \tag{4.12}$$

where the negative sign indicates the left hand side from ζ_0. Clearly, a small amount of light corresponding to α_2 is displaced in a much larger distance from ζ_0 ($\zeta_2 >> \zeta_1$) since $\alpha_2 > \alpha_1$. Furthermore, since the number of photons received is directly proportional to the aperture size, the sub-image at ζ_2 has x/y times the photons in ζ_1. As a result, the mean centroid of the wavefront in Fig. 4.6 can then be computed by

$$\begin{aligned} \langle\zeta\rangle &= \frac{1}{2}\{-f_L \times (z/y) \times y \; + \; f_L \times (z/x) \times x\} \\ &= \zeta_0. \end{aligned} \tag{4.13}$$

Thus it is the mean slope (Eq. (4.11)), and not the rms slope which determines the centroid of the observed image. The important point is that a comparatively small amount of light, provided it is displaced far from the origin, can significantly affect the estimate of the centroid. This in turn affects the estimate of the wavefront slope, and as shown in section 4.3, the performance of the Shack-Hartmann sensor.

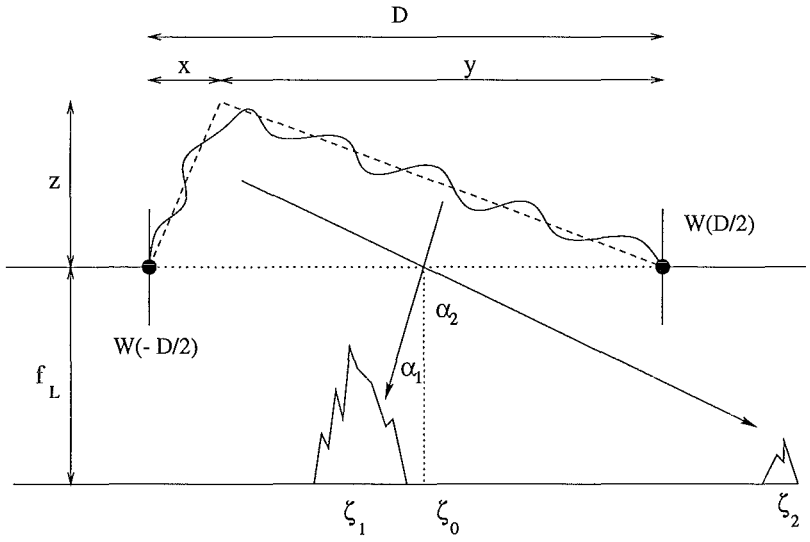


Figure 4.6: *Wavefront tilts that produce no centroid motion. The relatively small amount of light at ζ_2 averages the centroid of the brighter image (ζ_1) to the origin.*

4.2.2 The assessment of the wavefront distortion

The sensing of the wavefront distortion information is the first essential stage in compensating the disturbances associated with the atmospheric turbulence in real time. It is, however, difficult to measure the phase of the incoming wavefront at points across the telescope aperture directly. An alternative way is to measure the average wavefront slope over a set of regions in the aperture.

As noted in section 4.1.1, the wavefront can be sampled by an array of lenslets. Each lenslet defines a sub-aperture and cuts out a part of the wavefront. Each sub-aperture provides a low resolution sub-image of the object (hereafter called *spot*). When the spots are formed without the effect of the turbulence, they are located in the centroid of each sub-aperture (Fig. 4.7a). However, because of the presence of atmospheric turbulence, each spot is shifted inside the sub-aperture, proportional to the mean slope over the aperture. A diagram of the measurements that can be expected from the sensor is shown in Fig. 4.7b.

Mathematically, each sub-aperture gives two measurements, x_c and y_c. They are the displacements of the spot in two orthogonal directions denoted by x and y. It can be shown that [42]

$$x_c = \frac{\lambda f_L}{2\pi A_{sa}} \int\int_{sa} \frac{\partial \phi}{\partial u} \, du dv \tag{4.14}$$

(a) (b)

Figure 4.7: *Some typical spot measurements from the Shack-Hartmann sensor, (a) without the effect of turbulence, (b) with the presence of turbulence.*

and

$$y_c = \frac{\lambda f_L}{2\pi A_{sa}} \int\int_{sa} \frac{\partial \phi}{\partial v} \, \mathrm{d}u\mathrm{d}v \tag{4.15}$$

where A_{sa} is the area of the sub-aperture, $\frac{\partial \phi}{\partial u}$ is the u-derivative of the plane of the wavefront, and the integral is performed over the area of the sub-aperture. Rearranging Eqs. (4.14) and (4.15), the mean slope in x and y directions can be expressed as

$$\frac{1}{A_{sa}} \int\int_{sa} \frac{\partial \phi}{\partial u} \, \mathrm{d}u\mathrm{d}v = \frac{2\pi x_c}{\lambda f_L} \tag{4.16}$$

and

$$\frac{1}{A_{sa}} \int\int_{sa} \frac{\partial \phi}{\partial u} \, \mathrm{d}u\mathrm{d}v = \frac{2\pi y_c}{\lambda f_L}, \tag{4.17}$$

respectively. This shows in two dimensions that the mean slope of the wavefront over a sub-aperture can also be calculated from the displacements. From this information, an estimate of the wavefront can be made.

After obtaining a set of wavefront slope measurements (Fig. 4.7b), the wavefront itself has to be reconstructed from the measurements. A number of authors have addressed this problem [4, 56, 57]. Wavefront reconstruction can be divided into two main classes, namely the zonal and the modal approaches. When the wavefront is expressed in terms of the relationship of phase values over a small spatial area, or zone, the wavefront is said to be zonal. When the wavefront is expressed in terms of coefficients of polynomials extending over the entire pupil, it is said to be modal.

4.2.3 The estimation of the wavefront distortion

Southwell [57] has discussed three configurations that illustrate the grid positions for the x and y slope measurements as shown in Fig. 4.8. Figure 4.8a measures both the x and y slopes at the same point. Figure 4.8b illustrates a situation where the x and y slopes are processed separately. Figure 4.8c shows coincident x and y slopes, but with a displaced phase grid. The problem is to relate the slope measurement data from the Shack-Hartmann sensor to the phases of the wavefront at the grid points. In the zonal method the recursive

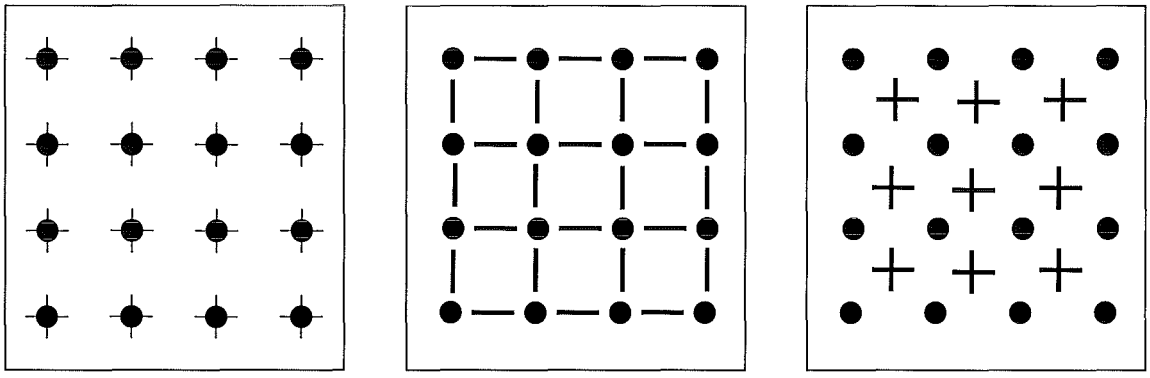


Figure 4.8: *Slope measurement sampling geometries and wavefront mesh points [57]. The horizontal dashed lines indicate position of x slope sampling. The vertical dashed lines are the y slope sampling positions. The dots are the estimated phase points.*

relationship of the slope measurements and the phases are given by [57],

$$\begin{aligned} \frac{S^x_{i+1,j} + S^x_{i,j}}{2} &= \frac{\phi_{i+1,j} - \phi_{i,j}}{p} \\ \frac{S^x_{i+1,j} + S^y_{i,j}}{2} &= \frac{\phi_{i,j+1} - \phi_{i,j}}{p} \end{aligned} \tag{4.18}$$

where S^x and S^y are the slope measurements in the x and y direction, respectively, $\phi_{i,j}$ is the phase value at the grid (i,j), and p is the distance between two phase measurements.

Considering a configuration with N^2 phase points and M slope measurements, this model is reduced to a single matrix equation:

$$\mathbf{S} = \mathbf{A}\Phi \tag{4.19}$$

where $\mathbf{A}$ is a $M \times N^2$ interaction matrix with its elements being $\pm(1/p)$ or 0, and Φ is a vector of length N^2 containing all the unknown phase values. Equation (4.19) can be

written in the following vector-matrix form

$$\begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_M \end{bmatrix} = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,N^2} \\ A_{2,1} & & & \\ \vdots & & \ddots & \\ A_{M,1} & & \cdots & A_{M,N^2} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{N^2} \end{bmatrix}. \tag{4.20}$$

The phase reconstruction is then achieved by minimizing the measurement error

$$\epsilon_s = \| \mathbf{S} - \mathbf{A}\mathbf{\Phi} \|^2 \tag{4.21}$$

where $\| \quad \|^2$ is the norm of a vector. A straightforward solution is to apply least-squares techniques [53]

$$\mathbf{\Phi} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{S}. \tag{4.22}$$

However, this solution cannot be used because $\mathbf{A}^T\mathbf{A}$ is singular. As a result the singular value decomposition is used to determine the pseudo-inverse of $\mathbf{A}^T\mathbf{A}$. Using this method, the matrix $\mathbf{A}$ can be decomposed in space defined by orthonormal matrices $\mathbf{U}$ and $\mathbf{V}$ as [53]

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{4.23}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$, respectively. The matrix $\mathbf{\Sigma}$ is a diagonal matrix containing the singular values of $\mathbf{A}$ along its diagonal. Thus,

$$\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T \tag{4.24}$$

and its pseudo-inverse can be computed as

$$(\mathbf{A}^T\mathbf{A})^+ = \mathbf{V}^T\mathbf{\Sigma}^{-2}\mathbf{V} \tag{4.25}$$

where a superscript $^+$ denotes the pseudo-inverse.

In the modal estimate, on the other hand, the incident wavefront onto the telescope aperture is written in terms of a set of basis functions as described earlier in Eq. (3.39),

$$\phi(x,y) = \sum_j^{\infty} a_j \psi_j(x,y) \tag{4.26}$$

where $\phi(x,y)$ represents the unknown wavefront incident onto the telescope aperture, $\psi_j(x,y)$ is the j^{th} basis function, and a_j is the weighting attached to the j^{th} basis function. Recall from section 4.1.1 that the measurements from the Shack-Hartmann sensor are proportional

to the local slope, or the first derivative of the wavefront. Therefore, the measurements from the Shack-Hartmann sensor can be expressed as

$$\begin{aligned} \frac{\partial \phi}{\partial x_i} &= \sum_j a_j \frac{\partial \psi_j(x,y)}{\partial x_i} \\ \frac{\partial \phi}{\partial y_i} &= \sum_j a_j \frac{\partial \psi_j(x,y)}{\partial y_i}. \end{aligned} \tag{4.27}$$

Here again, the model can be reduced to a matrix equation

$$\mathbf{S} = \mathbf{A}\mathbf{a} \tag{4.28}$$

where $\mathbf{a}$ is a vector comprising the coefficients of the basis functions.

It can be seen that no matter which approach is used, zonal or modal, the estimation problem can still be rewritten in generalized matrix notation as

$$\mathbf{m} = \Theta\mathbf{a} \tag{4.29}$$

where $\mathbf{m}$ is a vector containing all the slope measurements from the Shack-Hartmann sensor, Θ is the interaction matrix (described in the next section), and $\mathbf{a}$ is a vector containing the unknown phase values in the zonal approach, and is a vector containing the expansion coefficients in the modal approach. This linearization greatly simplifies the wavefront estimation.

Southwell [57] has demonstrated that for three different types of slope measurement sampling geometry shown in Fig. 4.8, Fig. 4.8a produces the least phase error, and the differences between each corresponding noise measurement become smaller as the number of phase points increases for modal estimation. In contrast, for zonal estimation, the differences between the Fig. 4.8a and Fig. 4.8b geometries become smaller, but the large gap to Fig. 4.8c geometry appears to persist even at large number of phase points. He also shows that for modal estimation the number of modes considered can be adapted to a particular problem. Furthermore, the modal estimation is computationally easier and faster. For these reasons, only the modal approach is considered here.

4.2.4 Modal sensing by use of Zernike polynomials

It should be noted in Eq. (4.26) that the choice of any complete set of orthogonal basis functions should produce identical results, since the wavefront estimate should depend on

the observed data and not the assumed basis functions. As mentioned in section 3.2.3, a statistically independent set of basis functions to describe a wavefront aberrated by Kolmogorov turbulence are the Karhunen-Loeve functions because they are statistically independent. Unfortunately, it is not possible to calculate these functions analytically. Wang and Markay [40], however, noted that for a circular aperture the low order Karhunen-Loeve functions are closely approximated by the Zernike polynomials. Roddier [39] then showed how the Karhunen-Loeve functions can be expressed in terms of the Zernike functions by diagonalization of the Zernike covariance matrix. It has also been shown that the Zernike polynomials have convenient mathematical properties for the analysis of the wavefront aberrations. Therefore, Zernike polynomials are often used as the basis functions.

In order to determine the lower and upper limit of the polynomials, it should be noted that $Z_1(x, y)$ is constant over the aperture, and since the average phase across the aperture does not affect the image it can be left out of the basis function. In addition, it is not possible to use an infinite number of Zernike polynomials in computing the results of this thesis. Therefore, the expansion of the phase $\phi(x, y)$ over J Zernike polynomials takes the form

$$\phi(u, v) = \sum_{j=2}^{J} a_j Z_j(u, v), \tag{4.30}$$

and the gradients in Eq. (4.27) become

$$\begin{aligned} \frac{\partial \phi(u, v)}{\partial u} &= \sum_{j=2}^{J} a_j \frac{\partial Z_j(u, v)}{\partial u} \\ \frac{\partial \phi(u, v)}{\partial v} &= \sum_{j=2}^{J} a_j \frac{\partial Z_j(u, v)}{\partial v}. \end{aligned} \tag{4.31}$$

Also mentioned in section 3.2.3 that the number of modes J required to achieve a certain Strehl value depends strongly upon D/r_0 (cf. Eq. (3.52)). For example, if $D/r_0 = 40$, then approximately 1000 Zernike terms must be corrected to be able to reach the Marechal's criterion (section 3.1.2). However, if the D/r_0 is reduced to 8, the number of modes decreases significantly to about 100. It is found, however, in the experiments that the results presented in this thesis were not significantly affected by truncation of the series in Eq. (4.30) at J=300.

Recall the covariance of the coefficients of the Zernike polynomials, $\mathbf{C}_a$, defined in Eq. (3.53). This covariance matrix provides information about the strength of the various modes and about the correlation between modes. The elements of the $\mathbf{C}_a$ matrix have been calculated

by Noll [38] for index of refraction fluctuations characterized by Kolmogorov power spectral density, and the covariances of the first eight Zernike modes, excluding piston are given by

$$\begin{bmatrix} 0.4577 & 0 & 0 & 0 & 0 & 0 & -0.0144 & 0 \\ 0 & 0.4577 & 0 & 0 & 0 & -0.0144 & 0 & 0 \\ 0 & 0 & 0.0236 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0236 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0236 & 0 & 0 & 0 \\ 0 & -0.0144 & 0 & 0 & 0 & 0.0063 & 0 & 0 \\ -0.0144 & 0 & 0 & 0 & 0 & 0 & 0.0063 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0063 \end{bmatrix} \times \left(\frac{D}{r_0}\right)^{5/3}. \tag{4.32}$$

It is clear from Eq. (4.32) that $\mathbf{C}_a$ is a function of $(D/r_0)^{5/3}$.

With the above covariance of the Zernike polynomials, the total phase variance across the aperture can be estimated as [38]

$$\begin{aligned} \sigma_\phi^2 &= \mathrm{Tr}[\mathbf{C}_a] \\ &\approx 1.04(D/r_0)^{5/3} \end{aligned} \tag{4.33}$$

where Tr denotes the trace of a matrix, which is the sum of its diagonal elements. Note that for a square aperture the variance is approximately $1.31(D/r_0)^{5/3}$ [9].

Consider now the interaction matrix $\mathbf{\Theta}$ in Eq. (4.29). This matrix relates the Zernike polynomials to the mean slope across the lenslets. For a 1 meter telescope with a secondary mirror dimension of 20 cm (20% of D), the first five columns of $\mathbf{\Theta}$ for a 2×2 lenslet take the values:

$$\mathbf{\Theta} = \begin{bmatrix} 4 & 0 & -5.8808 & -4.1584 & -4.1584 & \cdots \\ 0 & 4 & -5.8808 & -4.1584 & 4.1584 & \\ 4 & 0 & -5.8808 & 4.1584 & -4.1584 & \\ 0 & 4 & 5.8808 & -4.1584 & -4.1584 & \\ 4 & 0 & 5.8808 & -4.1584 & 4.1584 & \\ 0 & 4 & -5.8808 & 4.1584 & 4.1584 & \\ 4 & 0 & 5.8808 & 4.1584 & 4.1584 & \\ 0 & 4 & 5.8808 & 4.1584 & -4.1584 & \end{bmatrix}. \tag{4.34}$$

The columns of the matrix $\mathbf{\Theta}$ represent the values of Zernike polynomials in two orthogonal directions corresponding to the mean slope across the four lenslets, with the piston term

being excluded. Consider the first column, for example, each lenslet has tip (Z_2) term in the x direction and nothing in the y direction. If M denotes the total number of measurements, which is twice the number of lenslets, the dimension of $\mathbf{\Theta}$ is $M \times J$. Since $\mathbf{\Theta}$ is a function of D and not r_0, there is an asymmetry in the system performance when D and r_0 change proportionately. As an example, if D increases with a factor of 2, then $\mathbf{\Theta}$ reduces by 1/2. By contrast, $\mathbf{\Theta}$ is not a function of r_0.

In order to describe the optimal reconstructor, it is useful to begin with the noise free (nf) measurement:

$$\mathbf{m}_{nf} = \mathbf{\Theta a} \tag{4.35}$$

where $\mathbf{m}_{nf}$ has a dimension of $M \times 1$, and $\mathbf{a}$ is a $J \times 1$ vector. In practice, $\mathbf{\Theta}$ and $\mathbf{m}_{nf}$ are available and the problem of estimating the wavefronts thus reduces to an algebraic problem, namely solving a system of linear equations. The wavefront can only be expanded exactly in terms of a set of infinite number of Zernike polynomials, but since there are only finite number of measurements from the Shack-Hartmann sensor, it is an under-determined problem. To solve Eq. (4.35), it is necessary that the number of coefficients that can be estimated be smaller than or equal to M. An approach to estimate these coefficients is thus to estimate only the first X coefficients in the expansion, where $X \leq M$. All other coefficients are then put to zero, which yields,

$$\hat{\mathbf{a}} = (\mathbf{\Theta}^T\mathbf{\Theta})^{-1}\mathbf{\Theta}^T\mathbf{m}_{nf} \tag{4.36}$$

where $\hat{\mathbf{a}}$ indicates an estimated quantity of $\mathbf{a}$.

The covariance of the noise free measurements can then be derived:

$$\langle \mathbf{m}_{nf}\mathbf{m}_{nf}^T \rangle = \langle \mathbf{\Theta a}(\mathbf{\Theta a})^T \rangle = \langle \mathbf{\Theta a a}^T\mathbf{\Theta}^T \rangle, \tag{4.37}$$

which can be simplified by noting that $\mathbf{\Theta}$ is a constant matrix, so that

$$\langle \mathbf{m}_{nf}\mathbf{m}_{nf}^T \rangle = \mathbf{\Theta}\langle \mathbf{a a}^T \rangle\mathbf{\Theta}^T = \mathbf{\Theta C}_a\mathbf{\Theta}^T. \tag{4.38}$$

It is interesting to note that the following observations occur for a single aperture in the Shack-Hartmann sensor. As mentioned before, if the dimension of the telescope D is doubled, the interaction matrix $\mathbf{\Theta}$ is halved. However, recalling that the covariance of the Zernike polynomials given in Eq. (4.32) is a function of $(D/r_0)^{5/3}$, it is obvious that for a twice aperture dimension the covariance increases by a factor $2^{5/3}$. Therefore, from Eq.

(4.38) it can be deduced that the signal decreases by: $2^{-2} \times 2^{5/3} = 2^{-1/3}$, which agrees with the theoretical prediction from Eq. (3.61).

Unfortunately, in real world measurements noise is often inevitable, with the photon noise and the readout noise being present. The following section discusses how the noise can be included in the analysis, and how the problem of a limited number of measurements can be overcome.

4.2.5 Optimal estimation in the presence of noise

This sub-section describes the analysis of optimal estimation in the Shack-Hartmann sensor in the presence of noise. The analysis is performed in terms of the mean slope of the wavefront in the aperture in the manner described in section 4.2.1. To maintain consistency throughout this chapter, the unit (rad/m) is used throughout to express the mean slope of the wavefront.

The actual measurement in the Shack-Hartmann sensor is given by

$$\mathbf{m} = \mathbf{\Theta a} + \mathbf{n} \tag{4.39}$$

where a $M \times 1$ vector $\mathbf{n}$ models the noise term. Denote the covariance matrix of the photon noise of the measurements $\mathbf{m}$ as $\mathbf{N}$. This covariance matrix is defined as

$$\mathbf{N} = \langle \mathbf{n}\mathbf{n}^T \rangle. \tag{4.40}$$

This is again measured in terms of variance of the wavefront slope in the aperture. It is a function of the number of photons and the lenslet size since this determines the physical size of the spot on the detector. Using Eqs. (4.38) and (4.40), the covariance of the observed measurements is equal to:

$$\langle \mathbf{m}\mathbf{m}^T \rangle = \mathbf{\Theta C}_a \mathbf{\Theta}^T + \mathbf{N}. \tag{4.41}$$

Having defined $\mathbf{C}_a$, $\mathbf{\Theta}$, and $\mathbf{N}$, it is now convenient to consider the optimal reconstructor $\mathbf{\Omega}$ which is used to estimate $\hat{\mathbf{a}}$ from the observed measurements $\mathbf{m}$:

$$\hat{\mathbf{a}} = \mathbf{\Omega m}. \tag{4.42}$$

A simple and straightforward solution to Eq. (4.42) is through the pseudo-inverse method. Again, it should be noted that the number of measurements from the Shack-Hartmann

sensor is finite, but the degree of freedom of the wavefront is infinite which also implies that the number of unknowns to be estimated is infinite causing the pseudo-inverse technique to be inapplicable. This is a typical under-determined system of equation where the number of equations is less than the number of unknowns. In order to solve any under-determined equations, some extra information need to be used to ensure the estimation problem becomes well-conditioned, that is the number of unknowns is less than or equal to the number of equations.

One way to incorporate some extra information in the calculation is through the conditional probability described in section 2.3. Wallner [9] and Solomon *et al* [75] have derived an approach to solve Eq. (4.42), which is based on the calculation of the maximum a posteriori (MAP) estimate for the Zernike expansion coefficients. The problem can be formulated as to find the estimate $\hat{\mathbf{a}}$ that is most likely given the observed data $\mathbf{m}$. Mathematically, it can be expressed as:

$$\max_{\hat{\mathbf{a}}} Pr\{\hat{\mathbf{a}}|\mathbf{m}\}. \tag{4.43}$$

Recalling the Bayes theorem (Eq. (2.58)), the posterior probability distribution of Eq. (4.43) can be written as a function of the likelihood function and the prior distribution as,

$$P_r\{\hat{\mathbf{a}}|\mathbf{m}\} = \frac{P_r\{\mathbf{m}|\hat{\mathbf{a}}\}P_r\{\hat{\mathbf{a}}\}}{P_r\{\mathbf{m}\}}. \tag{4.44}$$

Taking the logarithm of Eq. (4.44) yields

$$\log[P_r\{\hat{\mathbf{a}}|\mathbf{m}\}] = \log[P_r\{\mathbf{m}|\hat{\mathbf{a}}\}] + \log[P_r\{\hat{\mathbf{a}}\}] - \log[P_r\{\mathbf{m}\}]. \tag{4.45}$$

The final term on the right hand side can be discarded from further consideration since it is not a function of $\hat{\mathbf{a}}$.

Using the knowledge that the expansion coefficients a_j are zero mean Gaussian random variables (section 3.2.3), the prior distribution can be written as

$$P_r\{\hat{\mathbf{a}}\} = \exp\left[-\frac{1}{2}\hat{\mathbf{a}}^T\mathbf{C}_a^{-1}\hat{\mathbf{a}}\right]. \tag{4.46}$$

Also using the knowledge that the measurement error can be approximated by a Gaussian function, the likelihood function of the measurements can be written as,

$$P_r\{\mathbf{m}|\hat{\mathbf{a}}\} = \exp\left[-\frac{1}{2}\hat{\mathbf{n}}^T\mathbf{N}^{-1}\hat{\mathbf{n}}\right]. \tag{4.47}$$

Taking the log of Eqs. (4.46) and (4.47), and substituting the logs into Eq. (4.45) yields

$$\log[Pr\{\hat{\mathbf{a}}|\mathbf{m}\}] = -\frac{1}{2}\left(\hat{\mathbf{a}}^T\mathbf{C}_a^{-1}\hat{\mathbf{a}} + \hat{\mathbf{n}}^T\mathbf{N}^{-1}\hat{\mathbf{n}}\right). \tag{4.48}$$

Substituting Eq. (4.39) into Eq. (4.48) results in

$$\begin{aligned}\log[Pr\{\hat{\mathbf{a}}|\mathbf{m}\}] &= -\frac{1}{2}\left(\hat{\mathbf{a}}^T\mathbf{C}_a^{-1}\hat{\mathbf{a}} + (\boldsymbol{\Theta}\hat{\mathbf{a}} - \mathbf{m})^T\mathbf{N}^{-1}(\boldsymbol{\Theta}\hat{\mathbf{a}} - \mathbf{m})\right) \\ &= -\frac{1}{2}\left(\hat{\mathbf{a}}^T\mathbf{C}_a^{-1}\hat{\mathbf{a}} + \hat{\mathbf{a}}^T\boldsymbol{\Theta}^T\mathbf{N}^{-1}\mathbf{m} - \mathbf{m}^T\mathbf{N}^{-1}\boldsymbol{\Theta}\hat{\mathbf{a}} + \mathbf{m}^T\mathbf{m}\right). \quad (4.49)\end{aligned}$$

Taking the partial derivatives of Eq. (4.49) with respect to $\hat{\mathbf{a}}$, and equating it to zero,

$$\frac{\partial \log[Pr\{\hat{\mathbf{a}}|\mathbf{m}\}]}{\partial \hat{\mathbf{a}}} = \hat{\mathbf{a}}^T\mathbf{C}_a^{-1} + \hat{\mathbf{a}}^T\boldsymbol{\Theta}^T\mathbf{N}^{-1}\boldsymbol{\Theta} - \boldsymbol{\Theta}^T\mathbf{N}^{-1}\mathbf{m} = 0. \quad (4.50)$$

Therefore, the solution can be written as,

$$\hat{\mathbf{a}} = \boldsymbol{\Omega}\mathbf{m}, \quad (4.51)$$

where the optimal reconstructor $\boldsymbol{\Omega}$ is given by

$$\boldsymbol{\Omega} = (\boldsymbol{\Theta}^T\mathbf{N}^{-1}\boldsymbol{\Theta} + \mathbf{C}_a^{-1})^{-1}\boldsymbol{\Theta}^T\mathbf{N}^{-1}. \quad (4.52)$$

Law and Lane [59] have shown that the optimal reconstructor given in Eq. (4.52) is the same as

$$\boldsymbol{\Omega} = \mathbf{C}_a\boldsymbol{\Theta}^T(\boldsymbol{\Theta}\mathbf{C}_a\boldsymbol{\Theta}^T + \mathbf{N})^{-1}, \quad (4.53)$$

which was originally derived in another form by Wallner [9]. The form in Eq. (4.53) has the advantage of lower computational complexity since there is only one matrix inversion, and the dimensions of $\mathbf{N}$ are much less than $\mathbf{C}_a$.

As a point of interest, if $\mathbf{N}$ becomes very large, that is when the noise on the measurement increases, the $\mathbf{N}$ matrix in Eq. (4.53) becomes dominant. This has a consequence that only the large eigenvalues of $\mathbf{C}_a$ contribute to the inverse in Eq. (4.53), and therefore the solution is mostly determined by the sum of low order Zernikes. As can be shown in Fig. 3.12, the lowest significant orders Zernikes are the tip and tilt, and therefore when the data is very noisy only the slope of the wavefront is estimated. Furthermore, if the data is very noisy it can be seen from Eq. (4.52) that the optimal reconstructor $\boldsymbol{\Omega}$, and therefore $\hat{\mathbf{a}}$ tend to go to zero, which implies that no corrections are performed. This is supported by the argument that it is useful not to perform any corrections for very noisy data since they may in fact contribute to the noise. On the other hand, when the noise level is very low, the optimal reconstructor is dominated by the interaction matrix which is a function of the wavefront sensor design.

If the lowest order aberrations in the random wavefront are corrected, one is interested in knowing how much wavefront distortion remains. This question is usually referred to

as the mean square *residual error*, Δ_J, as has been described in section 3.2.3. Since a_j was assumed to be Gaussian distributed, it is straightforward to argue that Δ_J also has Gaussian statistics. Therefore, the only other parameters needed to fully characterize the residual error is the mean and correlation function. It is also mentioned before (section 3.2.3) that a_j is assumed to be zero mean. This zero mean assumption and the linearity of the relationship between a_j and Δ_J, imply that Δ_J is also zero mean. This fact leaves only the requirement to specify the correlation function for the residual error in the Zernike coefficients as

$$\begin{aligned}\Delta_J &= \langle(\mathbf{a}-\hat{\mathbf{a}})(\mathbf{a}-\hat{\mathbf{a}})^T\rangle \\ &= \langle(\mathbf{a}-\mathbf{\Omega m})(\mathbf{a}-\mathbf{\Omega m})^T\rangle. \end{aligned} \tag{4.54}$$

Substituting Eq. (4.39) into Eq. (4.54), yields

$$\begin{aligned}\Delta_J &= \langle\{(\mathbf{I}-\mathbf{\Omega\Theta})\mathbf{a}-\mathbf{\Omega n}\}\{(\mathbf{I}-\mathbf{\Omega\Theta})\mathbf{a}-\mathbf{\Omega n}\}^T\rangle \\ &= \langle(\mathbf{I}-\mathbf{\Omega\Theta})\mathbf{a}\mathbf{a}^T(\mathbf{I}-\mathbf{\Omega\Theta})^T-(\mathbf{I}-\mathbf{\Omega\Theta})\mathbf{a}\mathbf{n}^T\mathbf{\Omega}^T-\mathbf{\Omega}\mathbf{n}\mathbf{a}^T(\mathbf{I}-\mathbf{\Omega\Theta})\mathbf{a}\mathbf{n}^T \\ &+ \mathbf{\Omega}\mathbf{n}\mathbf{n}^T\mathbf{\Omega}^T\rangle. \end{aligned} \tag{4.55}$$

Assume further that there is no correlation between the measurement errors and the signal, hence Eq. (4.55) becomes

$$\Delta_J = (\mathbf{I}-\mathbf{\Omega\Theta})\mathbf{C}_a(\mathbf{I}-\mathbf{\Omega\Theta})^T+\mathbf{\Omega N\Omega}^T. \tag{4.56}$$

It is now possible to evaluate the error in the reconstruction by simply taking the trace of Δ_J. A point of interest here is when there is no noise or when the noise level is negligible, the residual error only depends on the matrix $(\mathbf{I}-\mathbf{\Omega\Theta})$. This matrix in fact indicates how well $\mathbf{\Omega}$ inverts $\mathbf{\Theta}$. In the ideal case, the the product $\mathbf{\Omega\Theta}$ should equal the identity, and the residual error is then minimized to zero.

In contrast, when significant amount of noise is present, the reconstructor tends to go to zero as described earlier, reducing Eq. (4.56) to $\mathbf{C}_a$. This implies that the maximum error using this reconstructor is 1.04 $(D/r_0)^{5/3}$, in other words it is never worse than doing nothing. In practice, there are practical limitations that prevent the Shack-Hartmann sensor from achieving the ultimate performance as given in Eq. (4.56), which is now discussed.

4.2.6 Practical limitations of a Shack-Hartmann sensor

An ideal Shack-Hartmann centroid detector is one which includes infinitely fine sampling of the image, and would be free of any kind of noise. A large number of factors keep the sensor

from achieving this ideal performance, and they are generally referred to as sensor error. Sensor error is simply the difference between the actual wavefront tilt and the wavefront tilt seen by the sensor. This sensor error fundamentally limits the accuracy to which a distorted wavefront is estimated and consequently corrected.

The first factor that limits the performance of the sensor is associated with the destruction of information by centroiding the spots. The section of the wavefront cut out by a single lenslet is not flat, but it may have some perturbation on it in such a way that the corresponding spot is not smooth but is itself distorted. A Shack-Hartmann sensor generally uses only the centroid of each spot to determine the wavefront slope. Obviously, there is still appreciable information in such images beyond simply their centroid [74]. This loss of information by centroiding places limitations on the amount of sub-aperture information that may be extracted from a Shack-Hartmann image.

The sensor also requires sufficient photons to make accurate slope measurements. This is particularly critical in astronomy where the most scientific interest is usually associated with the faintest objects. In this case the light detection often fails to account for noise observed in the image. The physical origin of this noise is the quantized, random nature of photon interactions [1]. Specifically, the location of a photon's arrival in a detector can only be predicted statistically beforehand. This fundamental source of randomness gives rise to noise in the sensor which is often called photon noise, or also known as shot noise by some authors. Photon noise imposes limitations on the performance of the sensor which are generally more severe than the limitations imposed by turbulence effects alone, particularly under low light levels. Photon noise is a form of image-dependent noise, and the strength of the noise depends on the number and distribution of the detected photons. This type of noise is usually assumed to obey Poisson statistics described in section 2.3.2 [33].

When only a few photons are available, the ability to estimate the centroid of the wavefront slope is limited, which can be viewed as a fundamental limit as it is determined by the astronomical object of interest. This is best explained in the following simulations. A phase screen is generated using the algorithm described in Harding *et al* [37] assuming a turbulence of $D/r_0 = 4$ (see also section 3.2.3), and it is shown in Fig. 4.9a. Assuming incoherent imaging (section 3.1.3), a noise free speckle PSF is then generated using the diffraction theory described in section 3.1.2. The sum of the PSF is equal to unity, and it is shown in Fig. 4.9b. It can be viewed as the probability density function for the arrival

of a single photon in the image plane. A test image is used which consists of a point source with a maximum intensity of 1000 photons (Fig. 4.9c). This image is convolved with the PSF of Fig. 4.9b to produce the mean photon density in the image plane (Fig. 4.9d),

$$g(x,y) = \delta(x,y) \odot h(x,y). \tag{4.57}$$

The term $g(x,y)$ gives the mean number of photons expected at a pixel in the image plane. The actual number of photons received in a speckle image at each point (x,y) is denoted by $d(x,y)$ and modeled as an independent Poisson process with mean $g(x,y)$. Thus, the probability of receiving $d(x,y)$ photons at (x,y) is (cf. Eq. (2.40))

$$P_r\{d(x,y)\} = \frac{g(x,y)^{d(x,y)}\exp[-g(x,y)]}{d(x,y)!}, \tag{4.58}$$

where $g(x,y) \geq 0$.

Figure 4.9e shows the effect of image-dependent photon noise on the image shown in Fig. 4.9d. In this figure the presence of isolated photons located far away from the spot itself should be noted. This is not as obvious as a gray-scale display shown in Fig. 4.9f. These photons have a significant effect in centroid calculation (Eq. (4.13)), and have been neglected in previous analyses as discussed in the next section.

As the second factor, the finite number of samplings in the array detector is also a source of measurement error. The detector in the focal plane of a sub-aperture is segmented into a number of pixels. Obviously, if an infinite number of pixels was available, the sensor would achieve the ideal performance. Unfortunately, the size of a CCD array in pixels under the lenslets is a fixed quantity. In section 2.2 it was demonstrated the process of sampling a continuous function, and the choice of an appropriate sampling interval. Unless this sampling criterion is met, it will degrade the system's ability to sense and compensate for the high spatial frequencies of the wavefront distortion.

It is interesting to note, however, that increasing the sampling rate simply means that the total photons received by the detector are spread out over higher number of pixels, producing isolated photon events. As a consequence, the dynamic range drops proportionally to the sampling rate, a situation demonstrated in Fig. 4.10a. Note that the peak of Fig. 4.10a which corresponds to a higher sampling rate is now significantly lower (3) than that of Fig. 4.9d (50). The importance of isolated photons is readily apparent in Fig. 4.10b.

Third, the construction of a CCD provides another source of measurement error. The

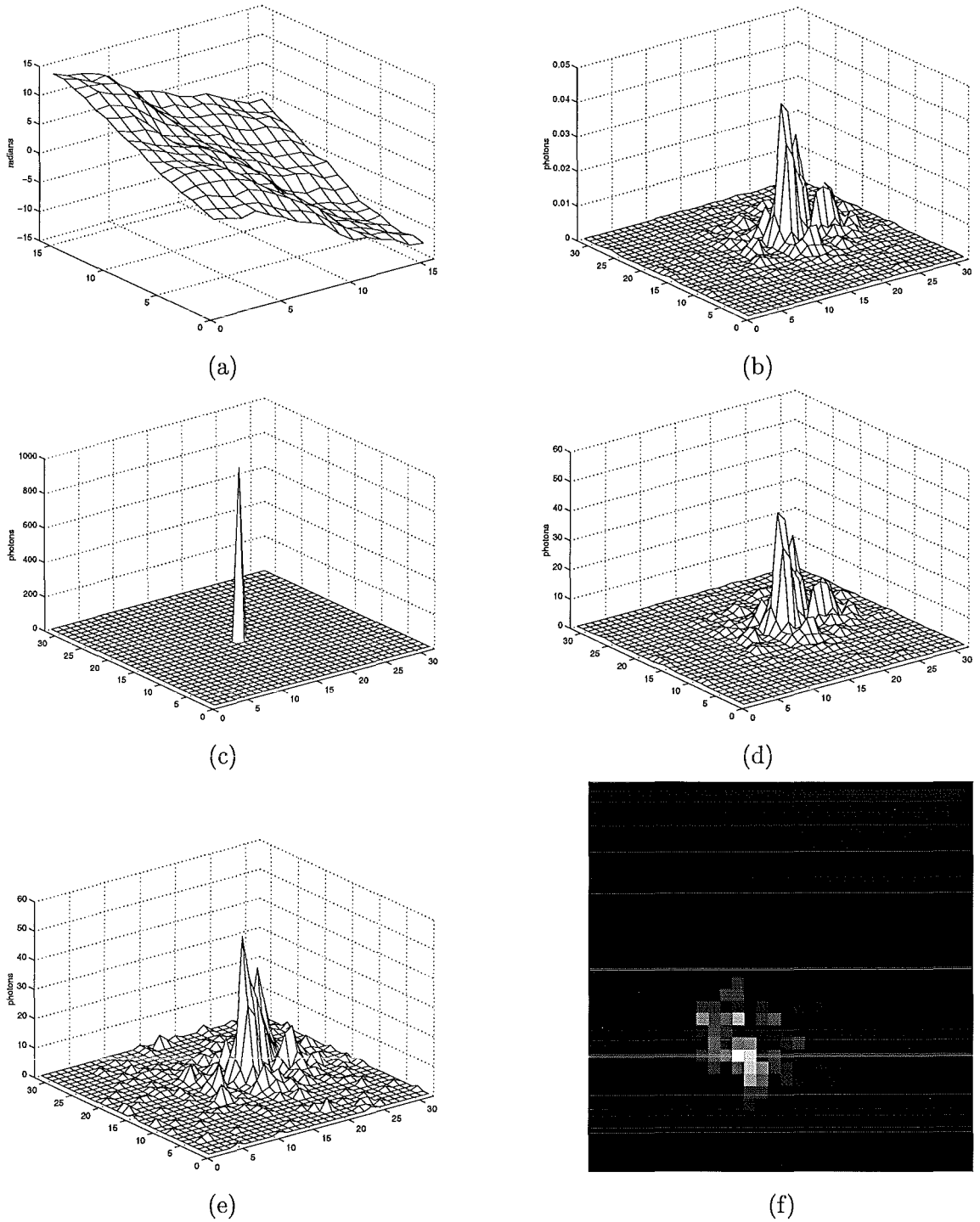
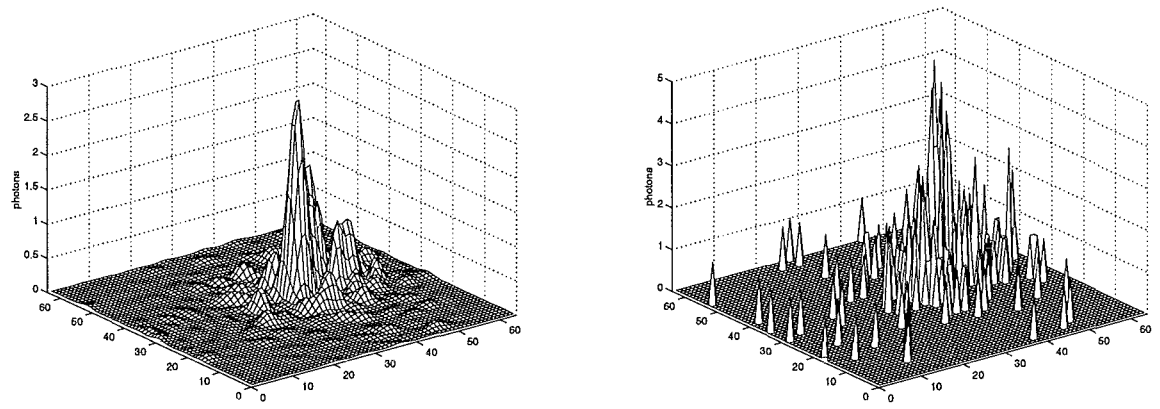


Figure 4.9: *(a) surface plot of a phase screen assuming $D/r_0 = 4$, (b) propagated phase screen onto an image plane, (c) test image of a point source, assuming the total photon count to equal 1000 photons, (d) distorted speckle image formed by convolving the point source with Fig. 4.9b, (e) noise-corrupted image due to image-dependent Poisson noise, (f) gray-scale plot of Fig. 4.9e.*



(a) (b)

Figure 4.10: *Increased sampling rate of Figs. 4.9d and 4.9e. Total photon count is still assumed to equal 1000, and $D/r_0 = 4$. (a) 64×64 speckle image, (b) noise-corrupted speckle image due to photon noise.*

surface of a CCD array in the lenslet focal plane is shown in Fig. 4.11. The pixels are assumed square and separated by a spacing known as dead space [44]. Shown also in Fig. 4.11 is a dashed circle which represents a possible location of the spot. The coordinates (ζ_i, ζ_j) are those of the centroid with respect to the origin at the center of the section of the CCD array. Error arises because of the finite size of the pixels and the dead space between pixels. A number of authors [44, 46] have studied the errors associated with this pixelization.

Fourth, the effects of additive noise further contribute to the degradation of images acquired with a CCD detector. The physical origin of this noise lies in the detector readout and amplification electronics [1]. Photons that fall into the active region of a pixel shown in Fig. 4.11 free electrons from the surface of the detector. Thus, the more light falling on the pixel, the more electrons that are liberated. Electrical fields then trap and isolate the mobile electrons in one pixel from those of other pixels until the electrons can be read out as an indication of how much light fell on the pixel. This process introduces a source of noise which is image-independent, and is generally referred to as readout noise [48]. Readout noise is characterized statistically as a Poisson random process (see section 2.3). Obviously, at very low light levels this type of noise becomes dominant, simply because there are not enough electrons to be liberated from incident photons. The readout noise increases with the number of pixels. Therefore, a trade-off should be made between the readout noise and the number of pixels.

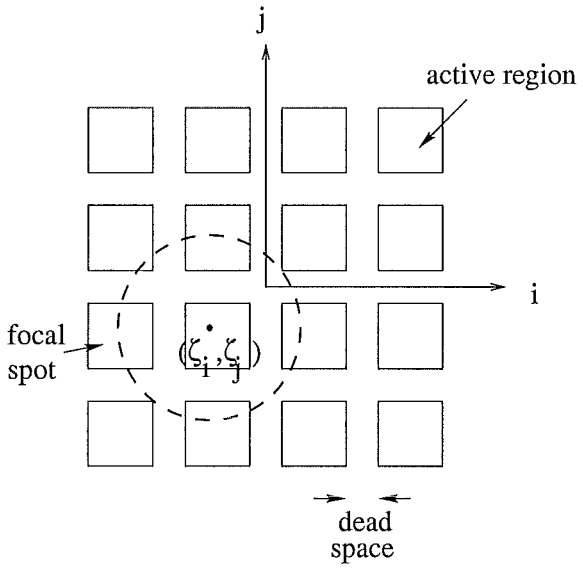


Figure 4.11: *Geometry of the focal spot and CCD array. The pixels or the active regions are assumed to be square.*

A number of different methods can be used to reduce the effects of readout noise. The simplest technique when the readout noise level is high is to use a quad-cell detector for each sub-aperture [18] since this minimizes the number of readout cells. The quad-cell detector consists of four individual detectors, with their edges parallel to the x and y axes as shown in Fig. 4.12. The performance of the quad-cell detector has been thoroughly analyzed in [19]. This technique, however, has serious drawbacks. If the spots do not fall precisely on the center of the quad-cell when no sub-aperture tilt is present, the tilt measurement is in error. Referring to Fig. 4.12, when the spot position moves, the position measurement accuracy changes [18]. When the spot falls completely into one side of the detector, the accuracy becomes very low, two situations depicted in the upper right and bottom left diagrams of Fig. 4.12, respectively. Furthermore, a quad-cell detector is limited in tilt dynamic range to ± 1 wave because the diffraction-limited spot has a radius equivalent to 1 wave of tilt [20]. For greater tilts, the spot moves fully onto one side or the other of the detector, and further motion does not change the response of the detector elements. This situation results in a low signal to noise ratio, and is shown in the bottom right diagram of Fig. 4.12.

Thus, whilst the quad-cell detector is indicated when readout noise is dominant, it is clearly sub-optimal. For this reason, the rest of the work in this thesis uses the light detector which is based on multi-pixel CCD described earlier, with the number of pixels being assumed to

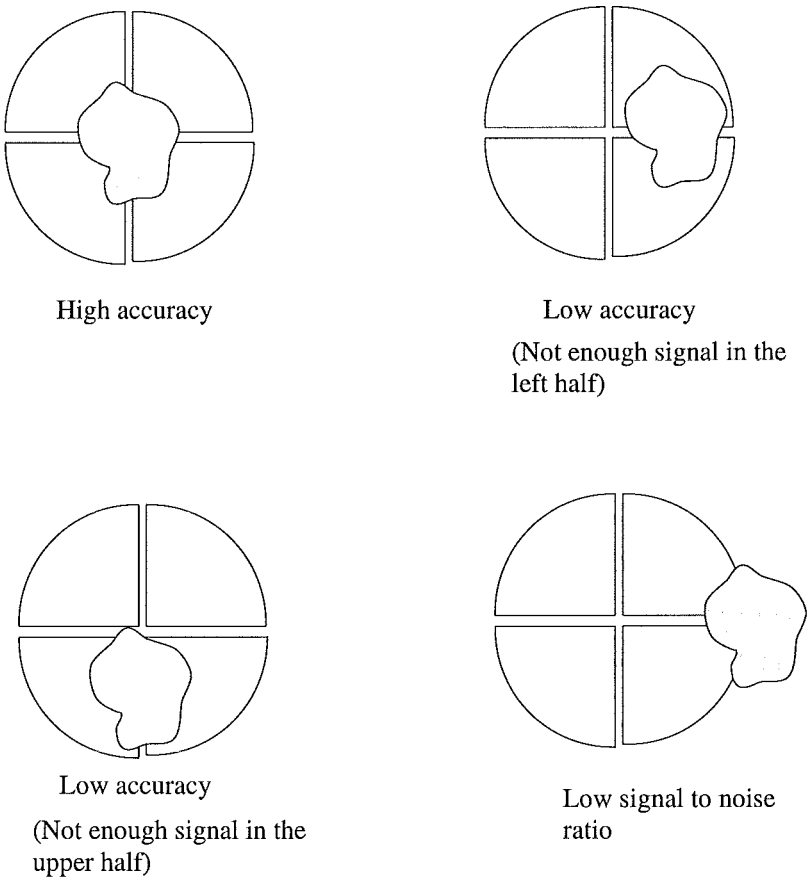


Figure 4.12: *Accuracy varies as the spot on the quad-cells moves. When the the displacement is very large, the spot moves completely off the detector and the measurement is lost.*

be large, and the readout noise is zero.

Finally, there is a factor associated with the time delay. Because the wavefront measurements are sampled due to the reading of the sensor, adaptive optics is a closed-loop system mixing continuous and sampled signals. The different elements in the system have been described earlier in section 1.2.2. The main part of the wavefront sensor is the detector which provides the data processed to obtain the wavefront measurements. Regardless of the type of the detector, a CCD or a quad-cell, it is characterized by its integration time. During this time, the detector accumulates photons coming from the science objects, but it also averages wavefront perturbations which are evolving at the same time [4]. Demerle *et al* in [4] have shown that for a wavefront sensor using a CCD, this time delay is equal to two sampling periods, where a sampling period is defined as 1/(the number of pixels). However, in the current state of the art technology, the effect of the time delay is well identified and

compensated.

4.3 Existing Analysis of SHWS Measurement Errors

It is well known [42,43] that the fundamental (i.e. non technological) limitation of the SHWS is that of centroiding spot accuracy. In this section an analysis technique is presented that is useful for evaluating and optimizing the performance of the sensor. The motivation of this detailed analysis is to obtain an ultimate performance of the sensor through the correct matrix of errors (**N**) in the sensor measurements. Sensor error, readout noise, pixelization, and the time delay effects described in the previous section are not considered here because they do not result in fundamental limitations, and it is not impossible to eliminate these deficiencies in the future technology. By contrast, insufficient photons is unavoidable when imaging distant objects. The theoretical analysis and simulation results herein prove useful for optimum design and application of the sensor.

4.3.1 Centroid variation (σ_ζ^2)

As mentioned in the previous section, in most practical situations the Shack-Hartmann measurements are corrupted by noise. Neglecting the additive readout noise from the detector, this noise is attributed to image-dependent photon noise alone, which can be modeled as being zero mean and independent of the sub-aperture wavefront slopes and also as independent from sub-aperture to sub-aperture. The evaluation of the slope measurement noise σ_α^2 depends merely on the photon noise predominating in the signal.

To analyze σ_α^2, it is necessary to consider the Shack-Hartmann sensor in more detail. Figure 4.13 illustrates this relationship of an incident tilted wavefront and a spot for a particular lenslet. Recalling Eqs. (4.16) and (4.17), the root mean square (rms) slope measurement noise σ_α is related to the rms of spot positional accuracy (hereafter called *centroid*), σ_ζ, by [42,43]

$$\sigma_\alpha = \frac{2\pi}{\lambda}\frac{\sigma_\zeta}{f_L} \quad [\text{rad/m}] \tag{4.59}$$

where $2\pi/\lambda$ is a constant that is usually referred to as the wave number k (cf. Eq. (3.6)). The rms of the positional accuracy of the spot is given by [43]:

$$\sigma_\zeta = \frac{\sigma_I}{\sqrt{n}} \quad [\text{m}] \tag{4.60}$$

where σ_I (in meters) is the rms width of the spot on the detector array surface (see Fig. 4.13), and n is the number of photons.

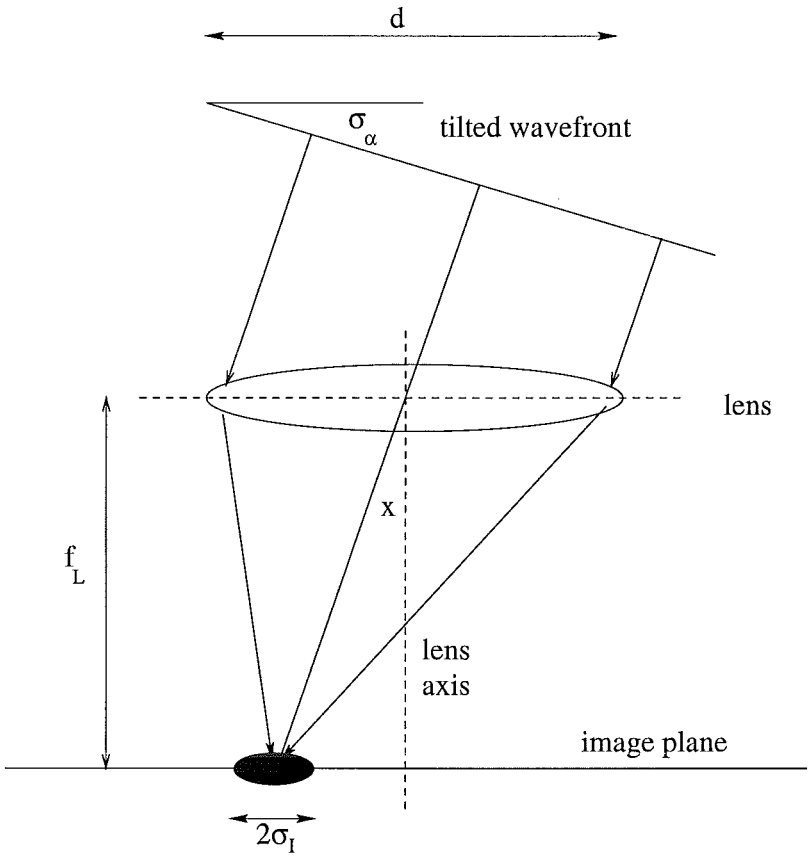


Figure 4.13: *Displacement of the spot on the detector due to the presence of atmospheric turbulence. The tilt of the wavefront is proportional to the displacement of the spot formed.*

Given the spot illumination $h(i,j)$ from each sub-aperture, the value of the centroiding along the i direction is simply (cf. Eq. (4.5)),

$$\zeta_i = \frac{\sum_i x_i \, h(i,j)}{\sum_i \sum_j h(i,j)} \tag{4.61}$$

where x_i is a vector representing the i axis. Since the variance of the measurement $h(i,j)$ is equal to the mean, $\langle h(i,j) \rangle$, the variance of the centroid owing to the Poisson noise in a Shack-Hartmann measurement in one direction can be derived using Eq. (2.45)

$$\sigma_\zeta^2 = \langle \zeta_i^2 \rangle - \langle \zeta_i \rangle^2$$

$$
\begin{aligned}
&= \frac{1}{\sum_i \sum_j h^2(i,j)} \left[\langle \sum_i x_i^2 \; h^2(i,j) \rangle - \sum_i x_i^2 \; \langle h(i,j) \rangle^2 \right] \\
&= \frac{1}{n} \sum_i x_i^2 \; \left[\langle h^2(i,j) \rangle - \langle h(i,j) \rangle^2 \right] \\
&= \frac{1}{n} \sum_i x_i^2 \; \langle h(i,j) \rangle. \qquad (4.62)
\end{aligned}
$$

As discussed in the previous section, the number of pixels in a sub-aperture is assumed to be large, and therefore the discrete sum can be replaced by continuous sum, hence the total variance in two orthogonal directions becomes

$$\sigma_\zeta^2 = \frac{2}{n} \int_{-\infty}^{\infty} x^2 \; \langle h(\mathbf{x}) \rangle \; \mathrm{d}\mathbf{x}. \qquad (4.63)$$

Note that Eq. (4.63) agrees with the theoretical derivation of the second order moment defined in Eq. (2.44).

4.3.2 Previous approximation of the spot

Many authors [42–44, 46] have used a simple Gaussian assumption for the illumination in the focal plane of the lenslet because it is mathematically tractable [46]. This assumption simplifies the integration into a complete analytical expression. Furthermore, the Gaussian

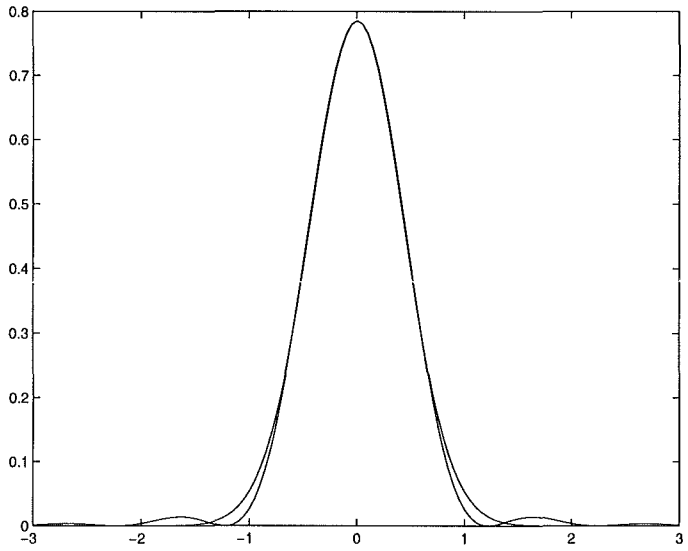


Figure 4.14: *Matching the Gaussian to the Airy pattern at the e^{-1} points. The Gaussian decays much faster to zero after the first minimum of the Airy disk (i.e. $3.83/\pi = 1.22$).*

distribution is matched to the e^{-1} points of the diffraction-limited spot as shown in Fig. 4.14. It should be noted that the function is normalized so

$$\int_{-\infty}^{\infty} h(\mathbf{x})\, \mathrm{d}\mathbf{x} = 1. \tag{4.64}$$

Welsh and Gardner [43] have shown that if the quantity r_0 is smaller than the lens size d, then the Gaussian distribution is matched to a diffraction-limited spot from a circular sub-aperture of diameter r_0. Recall from Eq. (3.16) that the minimum radius of the Airy disk allowing two images to be resolved is determined by the first minimum of the Airy disk (i.e. 3.83) divided by π, which is equal to 1.22. Since the spot is assumed to have a Gaussian profile with a e^{-1} point, a further division by e^{-1} (or $2\sqrt{2}$) is necessary to compute the rms width of the spot in radians. On the other hand, if $r_0 \geq d$ the Gaussian distribution is matched to a diffraction-limited spot from a square sub-aperture of side d and the first minimum now occurs at 1.05. Matching to the e^{-1} points as before results in a slightly narrower Gaussian approximation. In a closed analytical form, the standard deviation of the Gaussian, and therefore the rms width of the spot is then given in terms of r_0 and d by [43]

$$\sigma_I = \begin{cases} \frac{1.22\lambda f_L}{2\sqrt{2}r_0} & d > r_0 \\ & \text{[m]} \\ \frac{1.05\lambda f_L}{2\sqrt{2}d} & d \leq r_0. \end{cases} \tag{4.65}$$

Substituting Eqs. (4.60) and (4.65) into Eq. (4.59) gives [43]

$$\sigma_\alpha = \begin{cases} \frac{0.86\pi}{\sqrt{n}r_0} & d > r_0 \\ & \text{[rad/m]} \\ \frac{0.74\pi}{\sqrt{n}d} & d \leq r_0. \end{cases} \tag{4.66}$$

In the previously published work [43, 59, 75], the entries in the $\mathbf{N}$ matrix are the square of the rms slope measurement noise given in Eq. (4.66). The correlation between measurement errors is assumed to be zero and thus $\mathbf{N}$ is a diagonal matrix. The significance of $\mathbf{N}$ and its dependence on the lenslet (i.e. sub-aperture) size d for a fixed number of photons over an aperture is obvious in Eq. (4.66).

It is apparent in Eq. (4.66) that the measurement noise power σ_α^2 is inversely proportional to n. As n decreases, the effects of the slope measurement noise become very large producing large values in $\mathbf{N}$. As can be seen in Eq. (4.56), a large $\mathbf{N}$ also results in a high Δ_J. Figure 4.15 confirms this argument by showing the residual error defined in Eq. (4.56) for

$J = 300$ and $d/r_0 = 1$ as a function of the total number of photons and for two different sub-aperture sizes; 2×2 and 4×4, respectively. Note that when only few photons are available, the residual errors of a 4×4 lenslet array are higher than those of the 2×2 array. This situation does not last long, however, as can be seen in Fig. 4.15 that from approximately 100 photons the 4×4 array begins to perform better, despite having a larger D/r_0. This shows that the optimum sub-aperture size depends on the light level.

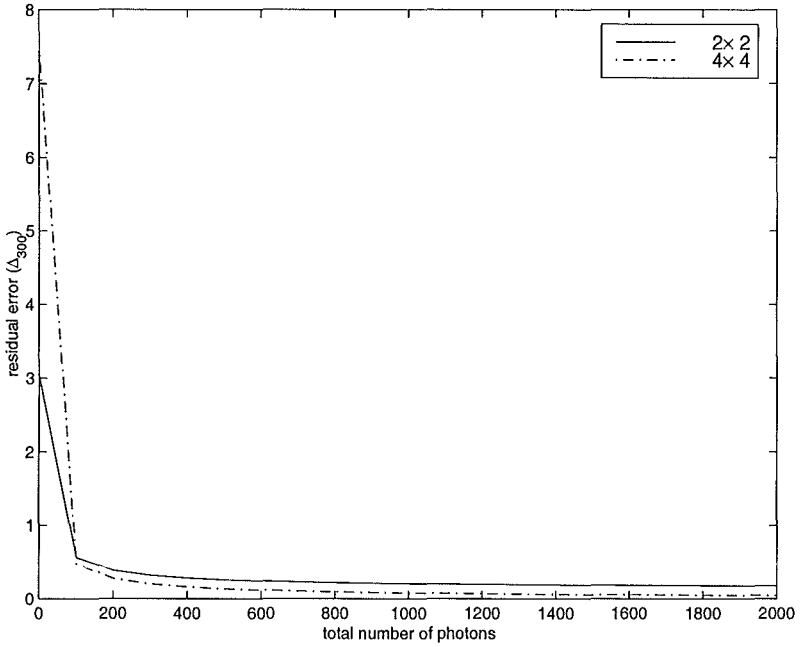


Figure 4.15: *The residual error ($Tr\{\Delta_{300}\}$) for 2×2 and and 4×4 lenslet array omputed using Eq. (4.56), as a function of the mean number of photons. It should be noted that for $d/r_0 = 1$, the overall system is $D/r_0 = 2$ for 2×2 and $D/r_0 = 4$ for 4×4.*

At $d > r_0$, the atmospheric turbulence becomes the limiting factor. As a consequence, the size of the spot remains the same, no matter how large the size of a lenslet is. However, recalling Eq. (3.61), as the sub-aperture size decreases by a factor 2 the signal level increases by $2^{1/3}$, and the noise variance associated with the photon noise increases by a factor of 4 since the number of photons is reduced by 4. On the other hand, at $d \leq r_0$, the sub-aperture becomes the band-limiting factor, and hence the size of the spot depends on the sub-aperture size. In this case, when the size of the sub-aperture is halved, the signal level still increases by $2^{1/3}$, and the noise variance is expected to be $4 \times 4 = 16$, where the first

4 comes from the number of photons, and the second is due to the fact that the size of the spot doubles.

It would appear from the preceding analysis that the lenslet size should be made as large as possible. However, as the sub-aperture size increases the number of sub-apertures and consequently measurements decreases. As a result, errors associated with finite number of measurements increase. In practice, an optimum sub-aperture size is found near $d = r_0$, since for lenslets larger than this the improvement in SNR resulting from the increased photons is compensated by the reduction in available measurements. As the sub-aperture size drops below r_0, the rapid increase in noise swamps the improvement due to extra measurements.

4.3.3 Use of actual instead of mean photon count

It is clear from the discussion presented in the previous section that the noise matrix $\mathbf{N}$ depends strongly on the number of photons received. The conventional Shack-Hartmann analysis relied simply on the mean number of photons for each sub-aperture [42, 43, 59]. Clearly more accurate results can be obtained if the actual number of photons is considered. This can be most clearly observed by noting that previously published work only assumes that there is always a measurement in each sub-aperture, although this may not be the case.

For example, if there is only 1 photon available for a 2×2 lenslet array, n in Eq. (4.66) takes the value of

$$\begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}.$$

If the actual number of photons of each sub-aperture is used in the calculation instead, typically only one sub-aperture that receives a photon and therefore n becomes, for example,

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Here, there is only one measurement out of four sub-apertures. As a result, using the actual photon count produces larger measurement errors since using the mean photon count implies that there is always a measurement in each sub-aperture. Using the mean photon count produces significantly lower measurement error at low light levels, and it thus over-estimates the performance of the sensor, which in fact should be much worse. Figure 4.16a shows the

residual errors of the 2 × 2 lenslet array for both cases. For clarity of the display, only a part of the graph is shown, which ranges from 100 to 1000 photons and the residual error from 0.2 to 1. For a comparison, Fig. 4.16b plots the residual error of the 4×4 lenslet array with the same range as Fig. 4.16a. It is readily seen from Fig. 4.16 that the residual error in the Shack-Hartmann sensor is in fact higher ($\approx$ 12%) than that is usually predicted if the actual number instead the mean number of photons of each sub-aperture is considered.

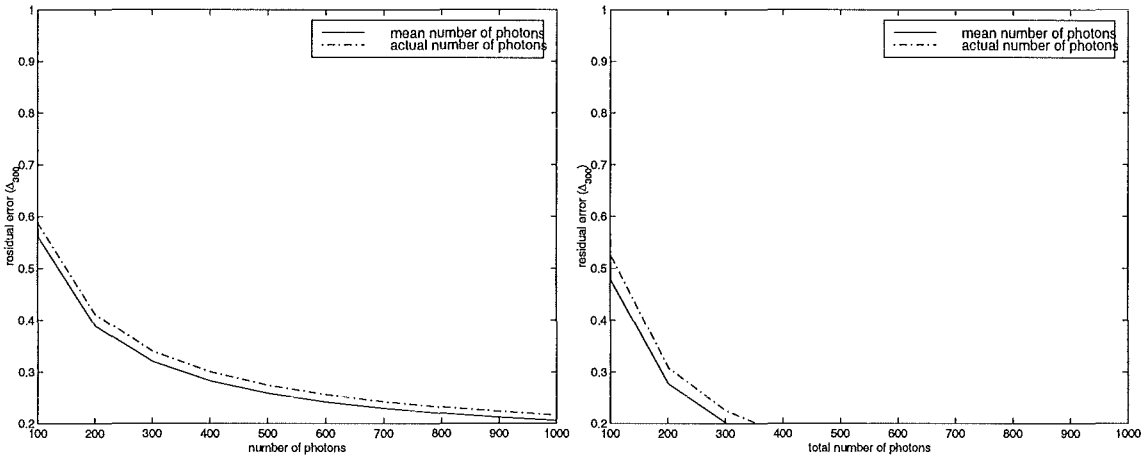


Figure 4.16: *Comparison of using the mean photon count and the actual photon count for (a) 2 × 2 lenslet array, and (b) 4 × 4 lenslet array.*

4.3.4 Long versus short exposure spot estimate

As noted in section 3.2.7 the difference between the long and short exposure images is in the tilt. A tilt of the incoming wavefront only shifts the center of the short exposure image and does not affect the image quality, whilst it blurs a long exposure image. A number of authors formulate the spot as being the Fourier transform of the long exposure OTF [42,43] ignoring the wavefront tilt over the sub-aperture. The calculations of the measurement errors so far given in Eqs. (4.65) and (4.66) are derived using the long exposure OTF.

Recently, it has been realized that it is more appropriate to base the analysis on the short exposure OTF [45,47]. Ellerbroek *et al* [45] obtained a result that was derived from complete knowledge of the short exposure case in their study of the performance characteristics of

the Shack-Hartman sensor. The result of the rms error in determining the location of the spot can be rewritten in terms of Eq. (4.65) as [45]:

$$\sigma_I = \begin{cases} \frac{1.22\lambda f_L}{2\sqrt{2}\ dK_1} & d > r_0 \\ & \text{[m]} \\ \frac{1.05\lambda f_L}{2\sqrt{2}\ dK_1} & d \leq r_0 \end{cases} \tag{4.67}$$

where

$$K_1 = \int_{-\infty}^{\infty} \langle \mathcal{H}(\mathbf{u}) \rangle_{SE} \mathbf{du}. \tag{4.68}$$

Note in Eq. (4.67) top assumes circular aperture, and bottom is for the square aperture. The corresponding slope measurement accuracy is

$$\sigma_\alpha = \begin{cases} \frac{0.86\pi}{\sqrt{n}\ dK_1} \times \frac{\pi}{4} & d > r_0 \\ & \text{[rad/m]} \\ \frac{0.74\pi}{\sqrt{n}\ dK_1} \times \frac{\pi}{4} & d \leq r_0 \end{cases} \tag{4.69}$$

where the factor of $\pi/4$ ensures that for an unaberrated instrument the standard deviation remains unchanged.

Recall from Eq. (3.70) that the integrals in K_1 represent the short exposure and long exposure resolution, respectively. Table 4.1 shows how K_1 varies with $d/r0$ based on the calculations presented in Fig. 3.17.

d/r_0	K_1
0	1
2	2.3615
4	3.4838
6	2.8263
8	2.3040
10	2.0510

Table 4.1: *The ratio of the short exposure to the long exposure resolution (K_1) as shown in Fig. 3.17.*

The difference between Eq. (4.67) and the expression given in Eq. (4.65) is in the denominator. As can be seen in Fig. 3.17, the resolution of the long exposure case approaches an

asymptotic limit fixed by r_0 which cannot be exceeded no matter how large d is. In contrast, the fact that significantly better resolution is obtainable with a short exposure was obvious, and therefore the noise is always lower when the short exposure case is considered. The point to be noted is that substantially higher resolution ($\approx 3.5\times$) can be achieved in the short exposure case when d/r_0 is nearly equal to 4 (table 4.1). This simply implies that the variance of the centroid of the spot achieves the minimum value at $d/r_0 \approx 4$, and consequently the noise is lowest at $d/r_0 \approx 4$. This also indicates that the use of lenslets with larger values of D/r_0 should further improve the SNR of the measurements.

Using Ellerbroek's result given in Eq. (4.69), the residual error is computed again and presented in Fig. 4.17. Figure 4.17 is a plot of the residual error, Δ_J, for $d/r_0 = 4$, and it should be compared with Fig. 4.15. It is readily apparent that the overall performance of 4×4 sub-apertures is always better than that of 2×2, despite having a larger D/r_0.

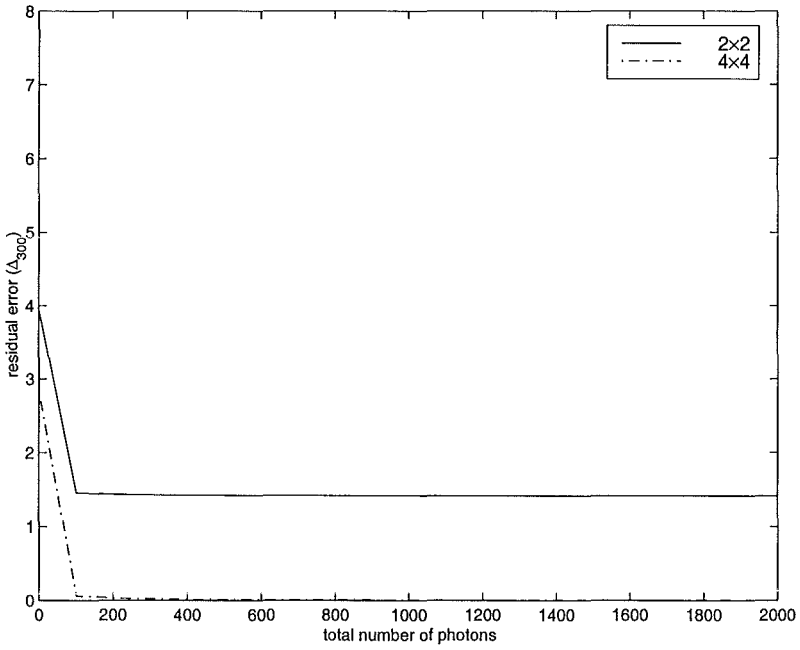


Figure 4.17: *The residual error ($Tr\{\Delta_{300}\}$) for 2×2 and 4×4 lenslet array computed using Eq. (4.69), as a function of the number of photons. Note that $d/r_0 = 4$ corresponds to $D/r_0 = 8$ for 2×2, and $D/r_0 = 16$ for 4×4.*

Using the short exposure spot as presented by Ellerbroek, and converting Eq. (4.63) into

rad^2, yields

$$\sigma_\zeta^2 = \frac{2}{n}\left(\frac{\lambda}{d}\right)^2 \int_\infty^\infty x^2 \, \langle h(\mathbf{x})\rangle_{SE} \, \mathrm{d}\mathbf{x} \qquad [\mathrm{rad}^2]. \tag{4.70}$$

The mean of the short exposure spot is related to its OTF by the two-dimensional inverse Fourier transform (section 2.2) as

$$\langle h(\mathbf{x})\rangle_{SE} = \int_{-\infty}^{\infty} \langle \mathcal{H}(\mathbf{u})\rangle_{SE} \times \exp[2i\pi\mathbf{x}\mathbf{u}] \, \mathrm{d}\mathbf{u}. \tag{4.71}$$

In cases where $\langle \mathcal{H}(\mathbf{u})\rangle_{SE}$ has circular symmetry, its transform $\langle h(\mathbf{x})\rangle_{SE}$ has also circular symmetry which can be solved by use of the Hankel transform of zero order (cf. Eq. (2.29)),

$$\langle h(r)\rangle_{SE} = 2\pi \int_0^1 \langle \mathcal{H}(\eta)\rangle_{SE} \times J_0(2\pi r\eta) \times \eta \, \mathrm{d}\eta \tag{4.72}$$

where r is the radial distance from the origin, and η is the radial spatial frequency variable.

Figure 4.18 illustrates a set of OTF curves for a normalized aperture dimension (cf. Fig. 3.16), where

$$\mathbf{u} = \frac{\lambda f_L \mathbf{f}}{d},$$

and $\mathbf{f}$ is a two-dimensional vector of spatial frequency in cycles/m. Each curve corresponds to $d/r_0 = 0$, 1, 2, and 4, respectively. The uppermost curve is the OTF of the perfect unaberrated system. The lowest curve is the OTF for the aberrated system with $d/r_0 = 4$. Figure 4.19 illustrates the corresponding one dimensional PSFs for the OTFs illustrated in Fig. 4.18 computed numerically from Eq. (4.72). The amplitudes of the PSF are plotted versus the dimensionless quantity r. The PSFs for both the unaberrated system and aberrated systems with d/r_0=1, 2, and 4, respectively, are shown. Recall from section 3.1.2 that the Strehl ratio compares the peak of the intensity of the PSF with that of the unaberrated, and therefore a Strehl of 1 indicates a perfect system. As expected, the Strehl ratio shows considerable degradation in performance as d/r_0 increases. Also the increased spreading of the PSF is apparent for larger d/r_0 values, and the significance of this is described in the next section.

4.4 New Analysis of the Measurement Errors

This section introduces a new analysis of the measurement errors in the Shack-Hartmann sensor. The aim is to obtain the correct matrix $\mathbf{N}$ by using the actual spot function in the noise computation. The use of Gaussian approximation, as extensively treated in the literature and described in section 4.3.2, does not yield the best fit to the spot, as

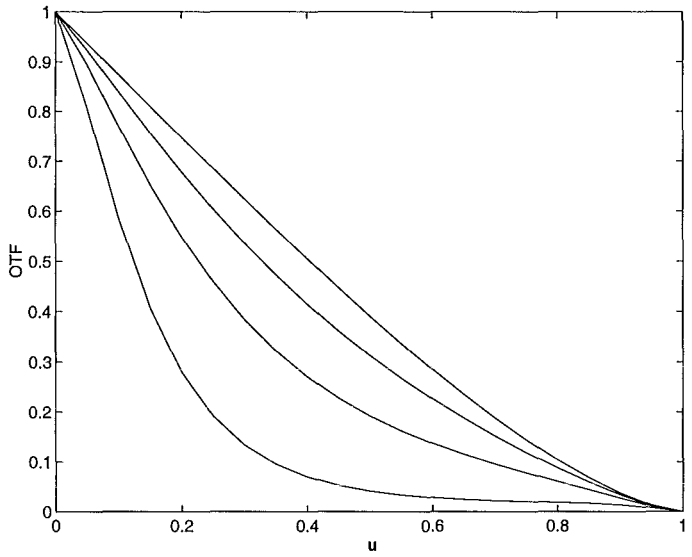


Figure 4.18: *OTFs for the $d/r_0 = 0$, 1, 2, and 4, respectively. The almost straight line corresponds to $d/r_0 = 0$, and the curve becomes narrower as d/r_0 increases.*

isolated photons located far away from the spot itself are not considered properly. Section 4.4.1 presents the theoretical analysis of the slope measurement noise without relying on the Gaussian approximation, and it is shown that the slope measurement variance depends critically on the detector size. In section 4.4.2 it is shown that the performance of the sensor is also affected by truncation owing to the finite size of the detector, a factor previously ignored. These two error metrics can then be used to compute the new entries for the noise matrix **N**, and determine the optimum size of the detector.

4.4.1 Theoretical analysis on slope measurement noise

A number of different methods can be used to measure the positions of the array of spots formed by a Shack-Hartmann sensor. Since we are only interested in the limits posed by the light level, readout noise is not considered and a finely sampled multi-pixel detector is analyzed. The problem as discussed in this section, is that tilt variance not only depends on the primary photon distribution (central region of the spot) but also any location where upon a photon falls.

In order to show why this is significant, it is convenient to analyze the diffraction-limited spot before discussing the aberrated system. The derivation of the diffraction limited spot

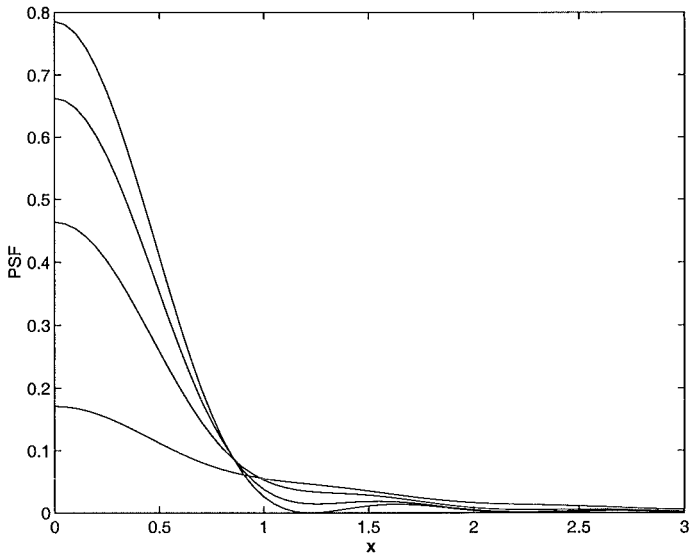


Figure 4.19: *PSFs for the $d/r_0 = 0$, 1, 2, and 4, respectively. The highest curve corresponds to $d/r_0 = 0$ and the lowest to $d/r_0 = 4$. The curve becomes wider and its peak drops as d/r_0 is increased.*

resulting in the well-known Airy function for a circular aperture is described in chapter 3, and is not repeated here. The Gaussian assumption to model the spot illumination described in section 4.3.2 decays much faster to zero than the diffraction-limited spot (see Fig. 4.14) outside the central region. Furthermore, no account is taken of the finite area of the detector. An accurate analysis is made by substituting Eq. (4.70) into Eq. (4.59), and defining a circular area on the detector surface, which results in

$$\sigma_{\alpha 1}^2 = \frac{2}{n}\left(\frac{2\pi}{d}\right)^2 \int_{A_\rho} x^2 \, \langle h(\mathbf{x})\rangle_{SE} \, \mathrm{d}\mathbf{x} \qquad [(\mathrm{rad/m})^2] \tag{4.73}$$

where A_ρ is the circular area on the detector with the radius:

$$\rho \geq \sqrt{x^2 + y^2}. \tag{4.74}$$

The spots are thus constrained to lie on a circle on the detector, whose ρ is allowed to increase gradually. This calculation has two significant differences from that of Welsh and Gardner [43]. First, the Gaussian function is used by Welsh and Gardner to represent $\langle h(\mathbf{x})\rangle_{SE}$ in Eq. (4.73), and second the size of the detector aperture is assumed to be infinite by Welsh and Gardner [43].

It should be noted that Eq. (4.69) is computed as $\rho \to \infty$ since there are no convergence difficulties when $\langle h(\mathbf{x})\rangle_{SE}$ is approximated by a Gaussian. Thus the limiting value is independent on the radius of the detector. If we assume a diffraction-limited spot (i.e. $r_0 \to \infty$)

in a circular aperture with $d = 25$ cm and $n = 1000$ photons, the variance derived by Welsh and Gardner is a constant value of

$$\sigma_\alpha^2 = \frac{2}{1000}\left(\frac{0.86\pi}{0.25}\right)^2 = 0.23 \qquad [(\text{rad/m})^2],$$

and is shown in dashed lines in Fig. 4.20.

Unlike the Gaussian approximation, the integral using the actual spot in Eq. (4.73) is analytically intractable, and it is necessary to perform the integration numerically. The numerical integration of Eq. (4.73) is given in Fig. 4.20 for $\rho = 0$ to $\rho = 5$, and it is shown in solid line. It is readily apparent that the variance computed using Eq. (4.73) depends critically on ρ, and increases without bound.

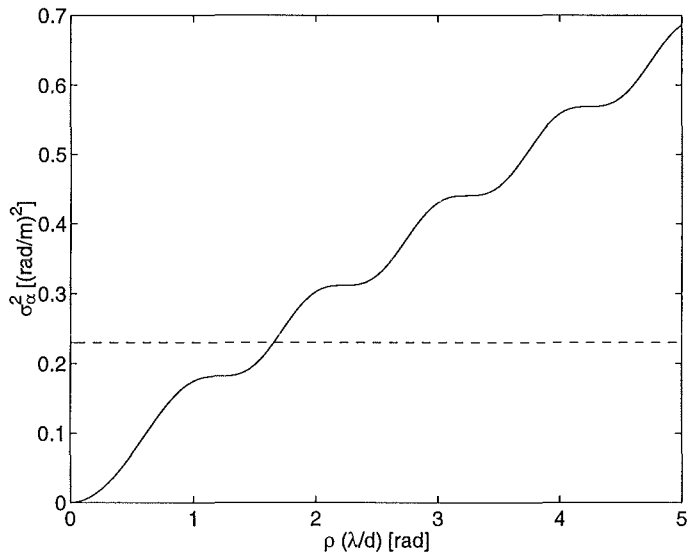


Figure 4.20: *The difference in the noise variance inherent in the centroid calculation using the Gaussian approximation (dashed lines) and the diffraction-limited spot (solid line), respectively. The sub-aperture size is set to $d = 25$ cm, and the number of photons $n = 1000$.*

The following intuitive argument can be used to explain the divergence of the above analysis. It can be seen in Fig. 4.18 that the OTF corresponding to the Airy disk possesses discontinuity in slope at $\mathbf{u} = 0$, and therefore an impulse in second derivative. This means that its Fourier transform behaves as $|\mathbf{x}|^{-2}$. Since the impulse remains in the second derivative the asymptotic decay of the diffraction-limited spot is once again proportional to $|\mathbf{x}|^{-2}$. Recalling Eq. (4.73), it can then be deduced that the term x^2 is cancelled out in the calcu-

lation, and thus produces a diverging function. This is even clearer when considering the diffraction-limited spot from a rectangular window with no turbulence, i.e.,

$$\langle h(\mathbf{x}) \rangle = \frac{\sin(x)\sin(y)}{xy}. \tag{4.75}$$

Substituting Eq. (4.75) into Eq. (4.73), and assuming a square integration region from $-\rho$ to ρ in both x and y directions, yields

$$\sigma_{\alpha 1}^2 = \frac{1}{n}\left(\frac{2\pi}{d}\right)^2 \int_{-\rho}^{\rho} x^2 \left(\frac{\sin(x)\sin(y)}{xy}\right)^2 \mathrm{d}x\mathrm{d}y. \tag{4.76}$$

This integral is of the form

$$K \int_{-\rho}^{\rho} \sin^2(\mathbf{x})\ \mathrm{d}\mathbf{x}$$

where K is a constant. This integral clearly diverges as $\rho \to \infty$.

Consider now the noise variance for spots produced by the aberrated systems with d/r_0 = 1, 2, and 4, respectively. Figure 4.21 illustrates a plot of $\sigma_{\alpha 1}^2$ versus the radius of the detector ρ. The lowest curve is the plot of $\sigma_{\alpha 1}^2$ for $d/r_0 = 1$, whilst the highest curve is that of the the aberrated system with $d/r_0 = 4$. It can be seen that the variances increase as ρ increases, and that the curves become steeper as d/r_0 is increased. Again, improved performance can be obtained by decreasing the sub-aperture size, d, relative to the Fried parameter r_0. This improvement, however, is gained at a reduction of photons in a given aperture. It can also be seen that the bumps in Fig. 4.20, corresponding to the rings of the Airy disk, become less obvious as the effect of turbulence increases.

4.4.2 Theoretical analysis on truncation error

The previous section demonstrates that although an ideal CCD array with infinite number of pixels was available, the performance of the sensor would still be limited owing to the photon noise which increases with the size of the CCD. Clearly, it is necessary to limit the size of the CCD to reduce this effect. This limitation, however, introduces another source of error due to the truncation. The truncation error simply occurs because the spot moves constantly at the image plane. In order to compute the error associated with this truncation, it is mathematically more convenient to hold the spot at the origin and allow the detector to move around. This situation is illustrated in Fig. 4.22. In this figure the spot is truncated by the finite area of the detector which is allowed to move at a variable distance m. The truncated region is denoted by the hatched area in the figure. The truncation thus causes displacement of the centroid from the origin, with the error becoming more severe

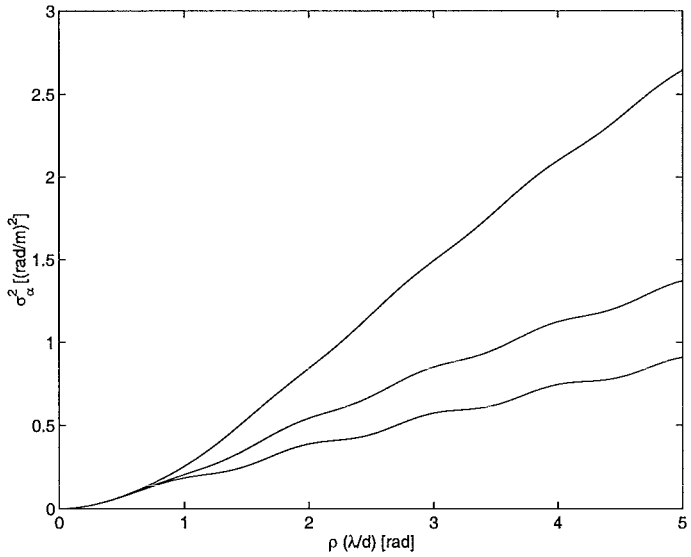


Figure 4.21: *Slope measurement noise due to the Poisson noise versus the radius of the detector using Eq. (4.73) with $d = 25$ cm, $n = 1000$ photons. $d/r_0 = 1$, 2, and 4, respectively. The lowest curve corresponds to $d/r_0 = 1$ and the highest corresponds to $d/r_0 = 4$. The diverging behaviour increases as d/r_0 is increased.*

as the detector moves away from the center of the spot. This bias is a function of the displacement, and is calculated in one direction by using the first moment as defined in Eq. (2.43):

$$b(m) = \frac{\int_{x=-\rho+m}^{\rho+m} \int_{y=0}^{\sqrt{\rho^2-(x-m)^2}} x\ h(x,y)\ \mathrm{d}x\mathrm{d}y}{\int_{x=-\rho+m}^{\rho+m} \int_{y=0}^{\sqrt{\rho^2-(x-m)^2}} h(x,y)\ \mathrm{d}x\mathrm{d}y} \tag{4.77}$$

where m is a variable to denote the horizontal shift of the detector, and the denominator represents the area of the spot inside the detector. The movement of the centroid is a Gaussian distribution (section 3.2.4) with a variance, σ^2, given by Eq. (3.61). The two-dimensional Gaussian pdf is therefore given by (cf. Eq. (2.39))

$$f_G(\mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x-\zeta_i)^2}{2\pi\sigma^2}\right] \exp\left[\frac{-(y-\zeta_j)^2}{2\pi\sigma^2}\right]. \tag{4.78}$$

Thus the expected variance due to the truncation in two-dimensional is

$$\sigma_t^2 = 2\left(\frac{2\pi}{d}\right)^2 \int_{-\infty}^{\infty} (b(m))^2 \times f_G(x)\ \mathrm{d}x \qquad [(\mathrm{rad/m})^2], \tag{4.79}$$

where the constant 2 accounts for the truncation effects in the two orthogonal directions.

Figure 4.23 plots the numerical integration given by Eq. (4.79) for $d/r_0 = 1, 2$, and 4. As can be seen in the figure, σ_t^2 decreases as the radius of the detector is increased. This is

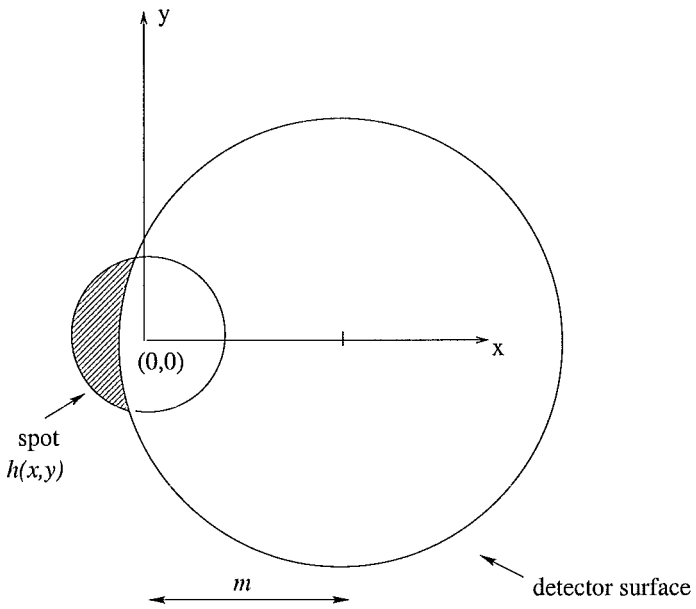


Figure 4.22: *Geometry of the error measurement on a detector array due to truncation. The spot is hold at the origin, and the detector is allowed to move at a variable distance m.*

not surprising since the larger the detector, the smaller the hatched area, and the less the effects of truncation.

Comparing Fig. 4.21 that increases with ρ against Fig. 4.23 that decreases, it indicates that there are intersection points which would represent the optimum radius of the detector. An intuitive argument is presented for why there should be an optimum size for a CCD array. Firstly, assume that a CCD array with a fixed size. Making the CCD array much larger than the spot size is not possible, because the photon noise diverges towards infinity. Thus the size of the CCD array should not be too large. On the other hand, if the size of the CCD array is small, then there may be some destruction of information due to the truncation of the short exposure image. Consequently, for a given CCD array size there should exist some intermediate size that balances both errors.

Figure 4.24 illustrates this point. For clarity of the display, only $d/r_0 = 1$ is considered with the number of photons being set to 10 and 100, respectively. It can be shown from Fig. 4.24 that the intersections representing where the Poisson noise exceeds the truncation noise are at 1.4 and 2.0 for 10 photons and 100 photons, respectively. However, the optimum radius

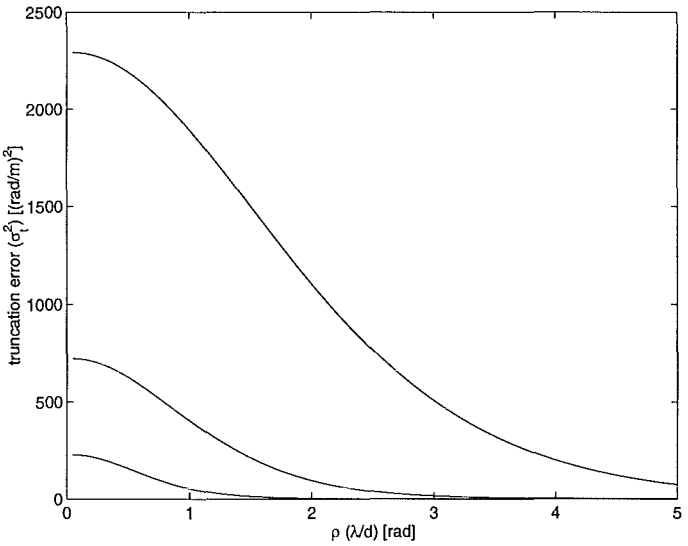


Figure 4.23: *Truncation error σ_t^2 computed using Eq. (4.79) versus the detector ρ which ranges from 1 to 5. The error achieves its maximum at the smallest radius and approaches zero at $\rho > 5$.*

is in fact determined by the minimum of

$$\sigma_\alpha^2 + \sigma_t^2, \tag{4.80}$$

which produces radius of 1.5 and 2.4. It is convenient for further reference to express the optimum radius in terms of the diffraction-limited spot width in radians. Recall from section 4.3.2 that the rms width of the diffraction-limited spot for a square aperture is $1.05/2\sqrt{2}$ radians. Denoting this value as β_{dl}, the optimum radius is now $\rho_{opt} = 3.2\beta_{dl}$ for 10 photons, and $\rho_{opt} = 6.5\beta_{dl}$ for 100 photons. Recalling Fig. 4.19 for $d/r_0 = 1$, it can be seen that these optimum values for the size of a detector still allow the spot to move around the detector, without significant truncation.

Table 4.2 tabulates the new entries of the noise matrix **N** using Eq. (4.80). In practical situation it is likely that **N** would be computed for different number of photons. Therefore, two variances corresponding to the optimum ρ are also given in table 4.2: one is for 10 photon counts and the other for 100 photons, and both are for $d/r_0 = 1$. It should be noted that for other values of d/r_0 and photon counts, the variance should be recalculated and a new graph should be obtained. It can be seen from table 4.2 that the lowest values (in italics) occur at $\rho_{opt} = 1.5$ (37) for 10 photons, and at $\rho_{opt} = 2.4$ (5.7) for 100 photons.

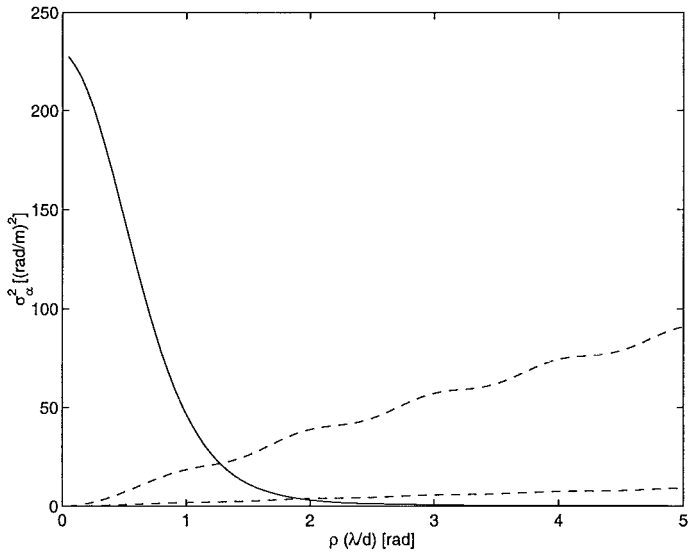


Figure 4.24: *Variance due to Poisson noise (dashed lines) and truncation effects (solid line). The number of photons is set to 10, and 100, respectively, and $d/r_0 = 1$. The intersections occur at $\rho = 1.5$ and $\rho = 2.4$ for 10 and 100 photons, respectively. See table 4.2 for details.*

4.5 Computer Simulations

In order to verify the analyses described in the previous section, computer simulations were carried out. It has been noted in the preceding analyses that there exist two types of error: one is due to photon noise and the other is due to the truncation. It is thus convenient to classify the simulation based on each type. The first two sub-sections present the simulation results for the above errors, and the last sub-section discusses the quantitative performance of the sensor in terms of the Strehl ratio.

The random midpoint displacement algorithm of Harding *et al* [37] is used to simulate the phase distortion that is due to Kolmogorov turbulence (see also section 3.2.3). The ratio of the lenslet diameter to Fried's parameter was set to $d/r_0 = 0$, 1, 2, and 4, respectively, which correspond to the earlier calculations (Fig. 4.21). The aperture dimension used in this simulation is 32 pixels inserted in an array of 128×128 zeros. The complex amplitude in the sub-aperture is then formed with the simulated phase and unit amplitudes. The Fourier transform of this complex array provides the spot intensity.

Radius	σ_α^2 [(rad/m)^2]	
ρ	$n = 100$	$n = 1000$
0.25	203.4	201.7
1	65.0	48.4
1.5	*37.0*	13.9
2	42.0	7.0
2.4	44.3	*5.7*
3	58.0	6.4
4	74.8	7.8
5	91.1	9.3

Table 4.2: *New entries of the noise matrix* **N** *for* $d/r_0 = 1$ *computed by a summation of the truncation error* (σ_t^2) *and the slope measurement noise* (σ_α^2)*. The numbers in italics represent the lowest variances and thus correspond to the optimum radius of the detector. The values should be referred to Fig. 4.24.*

4.5.1 Slope measurement error

First, consider the analysis given in Eq. (4.73) which results in diverging curves. It is convenient to define ρ_p as the radius used in the simulation, which is gradually increased from 1 to 20 pixels. The simulation is then carried out in the following steps:

- For each specified ρ_p, the spot with a predefined d/r_0 and n was first centered to produce a centroid at the origin. One reason to do this is to minimize any truncation effects due to the finite size of the detector. Another reason is to have a reference point at the image plane for the next calculations.
- The image was truncated and the truncated centroid $\bar{\zeta}_x$ and $\bar{\zeta}_y$ computed.
- The centered spot was then corrupted by image-dependent Poisson noise corresponding to 1000 photons. Note that the Poisson noise is necessary since it is often the best model to represent the image-dependent noise [61].
- The centroid displacement of the constrained noise-corrupted spot was computed in two orthogonal directions (ζ_x and ζ_y) with respect to the origin.
- The squared error was then defined as the sum of square differences in two orthogonal directions: $\epsilon_p = (\zeta_x - \bar{\zeta}_x)^2 + (\zeta_y - \bar{\zeta}_y)^2$.
- The above algorithm was repeated to produce 1000 different spots for every ρ_p.
- Finally, the variance of the centroid displacements was computed as $\sigma_p^2 = \langle \epsilon_p \rangle$.

To compare the centroid motion error of the simulation with the analysis shown in Fig. 4.21, it is necessary to have the same units. Therefore, these two following steps are performed. Consider first the ordinate which represents the values of σ_p^2. Since the simulation is done in pixels the obtained variances are in pixels^2. This should be converted into (rad/m)^2 to obtain the same dimension as Eq. (4.73). This conversion can be found by a multiplication by

$$\left(\frac{32}{128}\right)^2 \times \left(\frac{2\pi}{d}\right)^2 \tag{4.81}$$

where the ratio of the aperture dimension (32 pixels) to the array size (128 pixels) is to normalize the pixel dimension, while $(2\pi/d)^2$ is to convert further into (rad/m)^2.

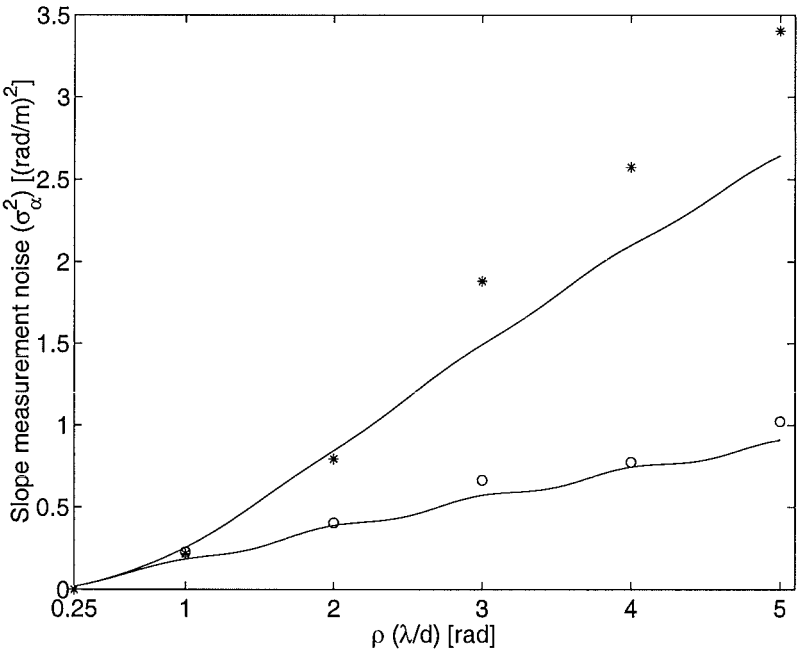


Figure 4.25: *Centroid variances due to Poisson noise for aberrated systems with $d/r_0 = 1$ (lower curve) and $d/r_0 = 4$ (upper curve). Circles (o) indicate the results of the simulations for $d/r_0 = 1$. Stars (*) correspond to the results of the simulations for $d/r_0 = 4$. The sub-aperture size d is assumed to be 25 cm, and $n = 1000$ photons. The theoretical values are drawn in solid lines.*

Second consider the abcsisca which represents the radius of the detector. Recall from section 2.2.7 that the sampling period at the detector plane must be less than $1/d$ to meet the Nyquist criterion. Since the aperture dimension is 32 pixels inserted in an array of 128 × 128 zeros, it is straightforward to deduce from Eq. (3.60) that the ρ_p should be divided by $\frac{128d}{32\lambda} = 4(d/\lambda)$ to have the same dimension as ρ. Figure 4.25 illustrates a plot of

σ_α^2 from both the analysis (Eq. (4.73)) and simulation results using the above configuration.

Radius ρ	σ_α^2 [(rad/m)^2]			
	$d/r_0 = 1$		$d/r_0 = 4$	
	analysis	simulation	analysis	simulation
0	0	0	0	0
1	0.2266	0.2281	0.2323	0.2171
2	0.4030	0.4046	0.8035	0.7935
3	0.6446	0.6641	1.8623	1.8813
4	0.7741	0.7747	2.4571	2.5740
5	0.9094	1.0230	2.6437	3.4030

Table 4.3: *Results for the new analysis of the SHWS Poisson noise by use of the short exposure spots. The total photon count is assumed to be 1000.*

In order to clarify the display of the results, only aberrated systems with $d/r_0 = 1$ and 4 are considered, rather than considering all the d/r_0 values addressed earlier. Note that the simulation results are very close to the theoretical values. Also, not surprisingly, the variances increase when ρ_p is increased, which again supports the analysis. Table 4.3 gives the values from which the graphs were drawn. As can be seen, the simulation results are slightly higher than those of the theoretical analysis. This difference is due to two main reasons: first, the integer-pixel accuracy in the centroiding algorithm used in this simulation contributes to the error particularly at small radius. As an illustration, consider a circle with a radius of 4 pixels on a 8×8 pixel detector array as shown in Fig. 4.26. It is readily apparent that a "smooth" circle cannot be achieved, which means that the accuracy of the measurements in the simulation is different from that of the theoretical analysis. In addition, a digital computer can only shift m pixels, where m is an integer, any shift that is less than a pixel introduces error. This error accumulates with the number of frames used in the simulation. Second, the finite number of sampling as described in section 4.2.6 is another source of error.

4.5.2 Truncation error

As with the previous simulation, the centroid of the spots is formed by integrating over a circle, whose radius is allowed to increase gradually. This time the spot is not centered

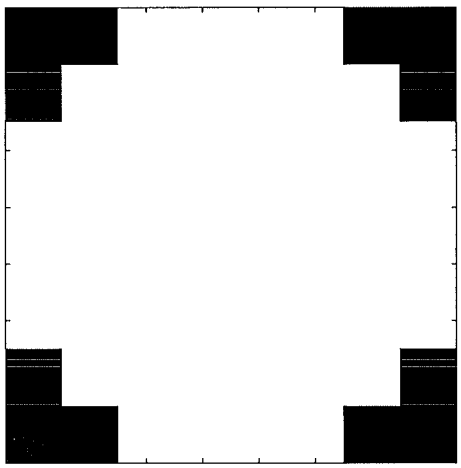


Figure 4.26: *The 8 × 8 pixel detector where upon a circle with a radius 4 is generated.*

before it is constrained to be within the circle and no noise is added. This produces a truncation of the spot, in which the severity of the truncation depends on both the position of the spot and the radius of the detector ρ_p. Recall from Fig. 4.22 that the further away the spot moves from the origin for a given size of the detector, the larger the area of the spot to be truncated. Similarly, as the size of ρ_p decreases for a given position of the spot, the truncated area also increases. The truncation error, as given analytically in Eq. (4.79) is simulated in the following steps:

- A spot was generated from a phase screen with a predefined d/r_0.
- The centroid displacement of the spot was computed in two orthogonal directions (ζ_x and ζ_y) with respect to the origin.
- A circular constraint with the radius ρ_p was then imposed on the spot.
- The centroid displacement of the constrained spot was computed in two orthogonal directions with respect to the origin, this time producing $\bar{\zeta}_x$ and $\bar{\zeta}_y$.
- The squared error was then defined as the sum of square differences in two orthogonal directions: $\epsilon_t = (\zeta_x - \bar{\zeta}_x)^2 + (\zeta_y - \bar{\zeta}_y)^2$.
- The above algorithm was repeated to produce 1000 different spots for every ρ_p.
- Finally, the truncation error was computed as $\sigma_t^2 = \langle \epsilon_t \rangle$.

Similar to the previous slope measurement simulation, σ_t^2 should be converted to $(\text{rad/m})^2$ before a comparison can be made. Using the same method described in the previous subsection, a multiplication of the ordinate by Eq. (4.81), and a division of the abcsisca by $4(d/\lambda)$ were carried out.

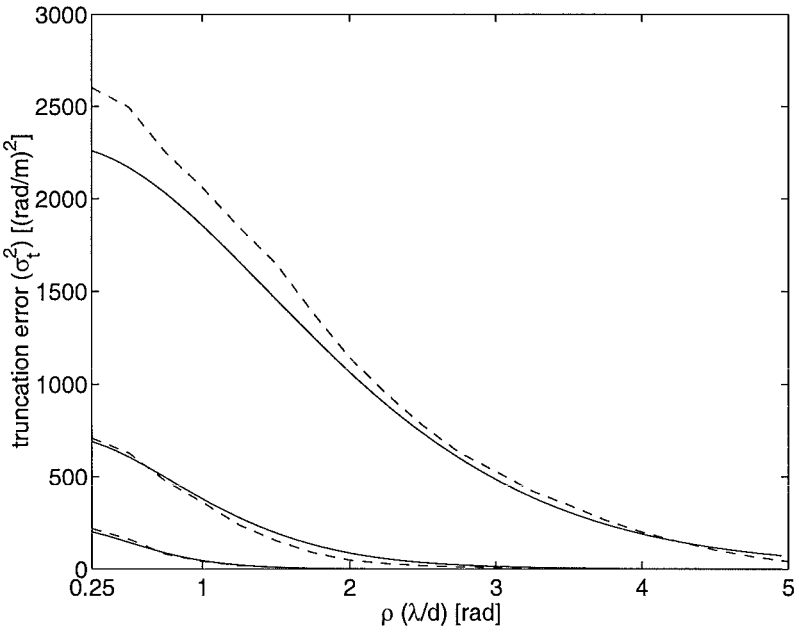


Figure 4.27: *Comparison of simulation and analytical result of truncation error, σ_t^2, as a function of the detector, ρ_p. d/r_0 is set to 1 (lower), 2 (middle), and 4 (upper).*

The results of the truncation error simulation described above is shown in dashed lines in Fig. 4.27, with d/r_0 being assumed to equal 1, 2, and 4. As a comparison, the truncation error obtained from the theoretical analysis (Fig. 4.23) is also shown in solid lines in Fig. 4.27. The radius of the detector aperture ρ_p is again increased from 1 to 20 pixels which correspond to $\rho = 0.25$ to $\rho = 5$.

It is apparent from Fig. 4.27 that increasing ρ_p reduces the truncation error, as theoretically predicted in Fig. 4.23. The values of Fig. 4.27 associated with the analytical results and simulation results, respectively, are listed in table 4.4. There are two points to be noted from this table. Firstly, as the radius of the detector tends to zero the centroid prediction also tends to zero. The truncation error variance thus approaches the variance of the spot motion predicted by Eq. (3.60). For $d/r_0 = 4$, for example, it is straightforward to compute the variance in two orthogonal directions using Eq. (3.61) in (rad/m)^2 as:

$$\begin{aligned} \sigma^2 &= \left(\frac{2\pi}{0.25}\right)^2 \times 0.36 \times 4^{5/3} \\ &= 2292. \end{aligned} \tag{4.82}$$

Secondly, the simulation results are consistent with those of analytical results. The fact that the simulation results are slightly higher than the analytical results are explained by

the same reasons described earlier in section 4.5.1.

Radius ρ	σ_t^2 [(rad/m)^2]					
	$d/r_0 = 1$		$d/r_0 = 2$		$d/r_0 = 4$	
	analysis	simulation	analysis	simulation	analysis	simulation
0.25	227.40	219.56	721.94	708.48	2292.00	2602.20
1	41.62	42.67	359.35	362.26	2060.72	2062.11
2	2.51	2.67	49.41	48.83	1140.17	1144.14
3	0.66	0.66	8.69	8.34	506.82	528.38
4	0.28	0.29	2.06	2.05	200.14	201.08
5	0.15	0.16	0.80	0.73	52.08	42.18

Table 4.4: *Results for truncation error computed analytically and by simulation. The numbers are obtained from Fig. 4.27.*

Consider now the optimal radius of the detector surface. Given Figs. 4.25 and 4.27, it is straightforward to extract the intersection points by superimposing these two figures. The two intersection points corresponding to $d/r_0 = 1$ and 10 photons and 100 photons are depicted in Fig. 4.28 as $\diamond$ and $\circ$, respectively. For 10 photons, the point where the Poisson noise exceeds the truncation noise is found to be 1.2 by simulation, whilst for 100 photons it is shown to be 1.9. These values are again supporting the analysis given in Fig. 4.24. Recall that the slope measurement noise power σ_α^2 is inversely proportional to n. As n increases, the effects of the slope measurement noise become negligible, and the optimal radius of the detector increases.

4.5.3 Quantitative performance evaluation

As a final evaluation on the performance of the Shack-Hartmann sensor, the Strehl ratio of the wavefront phase reconstruction is discussed. Recall that the Strehl ratio was related to the residual error in Eq. (3.50) as:

$$S = \exp[-\Delta_J] \tag{4.83}$$

where Δ_J is given by Eq. (4.56). It is interesting to compare the Strehls obtained from the conventional analysis of the sensor with those from the new analysis presented in table 4.2. The actual number of photons is used in this simulation, in which the number of photons per sub-aperture is assumed to vary from 1 to 100.

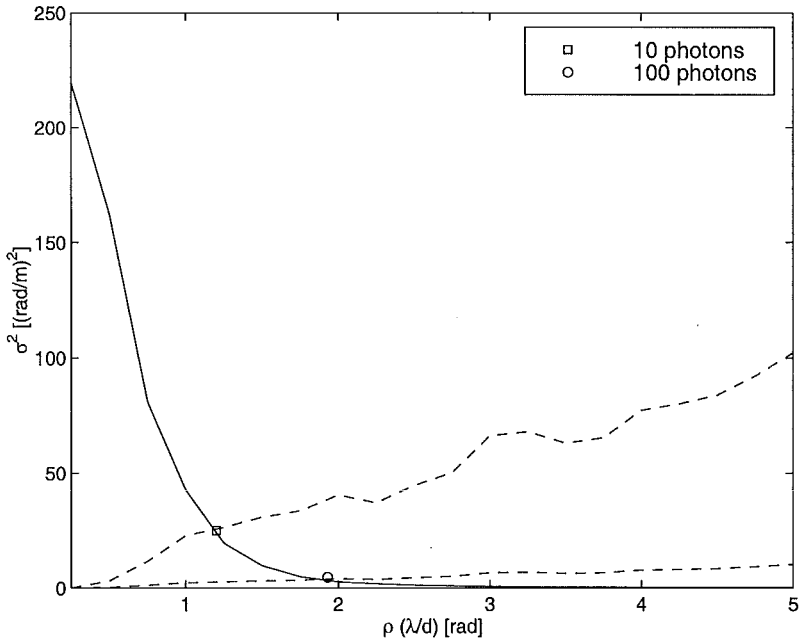


Figure 4.28: *Superimposing Fig. 4.27 on Fig. 4.21 for $d/r_0 = 1$. The optimum radius of 1.2 is found by simulation for 10 photons, and 1.9 for 100 photons.*

The results of such a calculation are presented in Fig. 4.29 for the following configuration. A telescope is assumed to have a diameter of 1 m, and the turbulence to be described by $r_0 = 0.25$ m, therefore the ratio D/r_0 is equal to 4. Consider first a 2×2 lenslet array, which produces 4 sub-apertures with each having a size of $d = 0.50$ m. The radius of the detector array is varied from $\rho = (\lambda/d)$ (dashed lines), $1.5(\lambda/d)$ (dash-dotted lines) to $2.4(\lambda/d)$ (crosses). The residual error using the result by Welsh and Gardner [43] is shown in solid line in Fig. 4.29, while those obtained from the new analysis presented in table 4.2 are drawn in dashed lines. Two significances are obvious in Fig. 4.29. First, the Strehls of the new analysis are significantly lower than those of the conventional analysis, although these values depend on the number of photons. Second, it can be seen from the new analysis that the curve goes up when the radius of the detector is increased from $\rho = (\lambda/d)$ to $\rho = 1.5(\lambda/d)$ before it goes down again at $\rho = 2.4(\lambda/d)$ for lower number of photons. This behaviour, however, changes slightly when the number of photons reaches 100, where $\rho = 2.4(\lambda/d)$ now produces highest Strehl. This is a consequence of the optimum size of the detector presented in the previous section.

In addition, the Strehls obtained from the conventional analysis reach a value around 0.52

when 100 photons/sub-aperture are available, while those obtained from the new analysis reach only around 0.4. The fact that the achieved Strehls are significantly lower when the new analysis is used partly explains the disagreements between theory and actual performance.

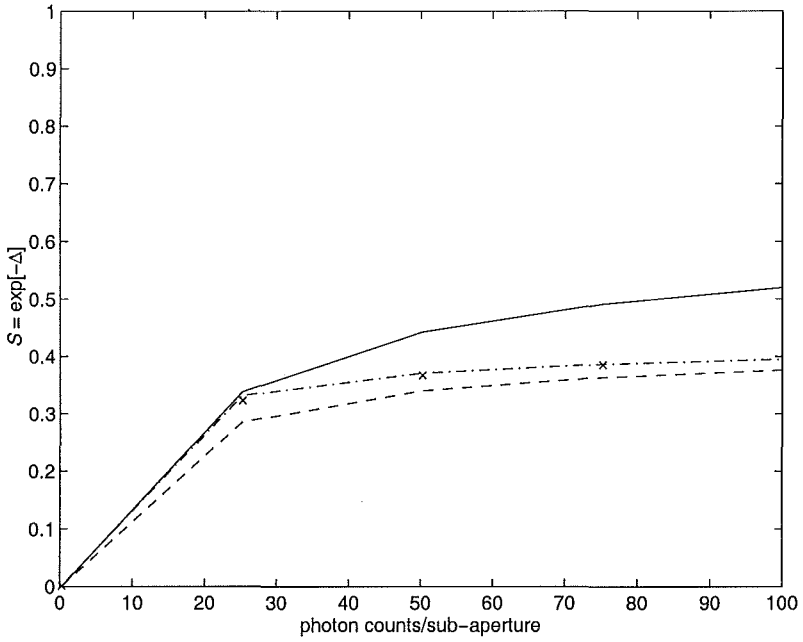


Figure 4.29: *Strehl ratio: a comparison between the conventional analysis (solid line) and the new analysis ($\rho = (\lambda/d)$ (dashed lines), $1.5(\lambda/d)$ (dash-dotted lines), $2.4(\lambda/d)$ (crosses)) for a 2×2 lenslet array and $D/r_0 = 4$. Table 4.2 is used to compute the latter.*

The same results for a 4 × 4 lenslet array are plotted in Fig. 4.30. In this case each sub-aperture has a size of $d = 0.25$ m. It is obvious that higher values of Strehl are achieved when more sub-apertures are used. This also confirms the residual error calculation shown in Fig. 4.15. When there are 100 photons/sub-apertures available the conventional Shack-Hartmann sensor achieves a Strehl of 0.72. However, using the new analysis it can be seen that for this number of photons the Strehl varies only from 0.53 to 0.56 depending on the radius of the detector.

As a final comment, note from Figs. 4.29 and 4.30 that the truncation error described in section 4.4.2 is readily apparent at radius $\rho = (\lambda/d)$ since the achieved Strehls are comparatively much lower than those of radius $1.5(\lambda/d)$ and $2.4(\lambda/d)$. This accounts for

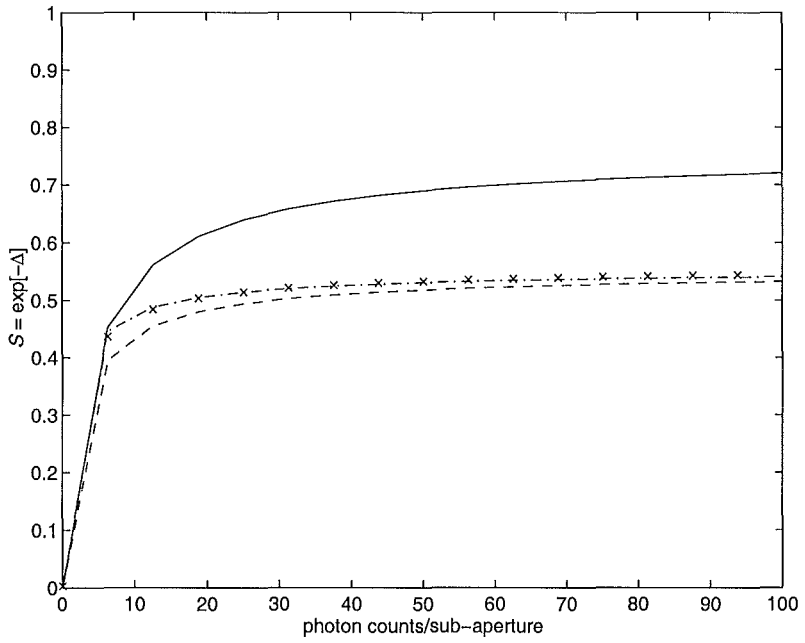


Figure 4.30: *Strehl ratio: a comparison between the conventional analysis (solid lines) and the new analysis ($\rho = (\lambda/d)$ (dashed lines), $1.5(\lambda/d)$ (dash-dotted lines), $2.4(\lambda/d)$ (crosses)) for a 4×4 lenslet array and $D/r_0 = 4$. Table 4.2 is used to compute the latter.*

the differences between the conventional analysis and the new analysis. As may be expected, these differences become smaller as the actual number of photons increases. This is because the slope measurement noise σ_α^2 is inversely proportional to n, and thus when n is very large, no matter which approximation is used for the spot, σ_α^2 should produce the same value.

4.6 Summary

The main contribution of this chapter is the presentation of a complete analysis of the most commonly used wavefront sensor in an adaptive optics system, namely the Shack-Hartmann sensor. Some practical limitations of the sensor are discussed. Along with these limitations centroiding accuracy of the sensor's measurements are presented that characterizes the variance of the additive slope measurement noise σ_α^2.

The slope measurement noise has been investigated by considering the actual short exposure spot in the focal plane. The detector is assumed to be a perfect multi-pixel CCD array.

Additive readout noise and effects due to finite pixel sizes have been ignored here since they are not fundamental measurement errors. Furthermore, the conventional Shack-Hartmann analysis relied simply on the mean number of photons for each sub-aperture. This is shown to be less accurate since it relied on the assumption that there is always a measurement in a sub-aperture, which may not be the case. It has been demonstrated that using the mean photon count produces significantly lower measurement error at low light levels, and thus the previous work has overestimated the performance of the sensor.

A new analysis is presented which is different from the previously published work in two main issues. First, the actual photon count is used in the sensor measurements. Second, the short exposure spot in the calculation of the measurement errors is used instead of the Gaussian approximation. From the latter, it has been shown that the derived variance due to the photon noise diverges as the size of the detector is increased. This differs from the previous measurement error calculation which showed no sensitivity to reasonable variations of the detector size. It has been discussed that it is necessary to limit the size of the detector to reduce the effect of the noise. However, this limitation introduces truncation error since the spot moves constantly at the image plane. These two errors are then used to determine the optimum radius of the detector.

The optimum detection radius, however, is quite sensitive to changes in the number of photons. It is shown, that as the number of photons increases, a larger detector must be used to achieve the optimum performance. The optimum radius is shown to be 1.5 for 10 photons, and 2.4 for 100 photons. These values can also be expressed in terms of the rms width of the diffraction-limited spot β_{dl} as: $3.2\beta_{dl}$ and $6.5\beta_{dl}$, respectively. Both cases are performed at a turbulence of $d/r_0 = 1$. Furthermore, based on these two errors, the new entries of the noise matrix $\mathbf{N}$ are also derived. This matrix is particularly important for computing the phase residual error in the Shack-Hartmann reconstructions.

Finally, using numerical simulations, it is shown that the analysis is quite accurate. The simulations also indicate that the sensor is quite sensitive to the truncation of the spot. The phase error of the Shack-Hartmann reconstructions is shown to be significantly higher when the new matrix $\mathbf{N}$ is used. In other words, the Strehls of the Shack-Hartmann reconstructions are lower when the actual short exposure spot is used in the calculation of the slope measurement noise instead of the Gaussian approximation as has been treated extensively in the literature. With $D/r_0 = 4$, a sub-aperture size $d = 0.25$ m, and 100 photons/sub-

aperture, the expected Strehl is now 0.5 as opposed to 0.7 in the conventional analysis. The numerical results also indicate that the Strehl ratio is quite sensitive to changes in the sub-aperture size and the number of photons.

Chapter 5

PHASE RETRIEVAL IN ASTRONOMY

In the previous chapter it was shown how a particular wavefront sensor, i.e. the Shack-Hartmann sensor can be used to estimate a wavefront phase. It was also shown that the accuracy of wavefront estimates is heavily dependent on the accuracy of the centroid estimates, and even in the absence of readout noise the performance is limited.

A promising alternative to wavefront sensing for adaptive optics system is a technique based on a direct phase retrieval algorithm. The most significant potential advantage of this approach is related to the fact that it uses the full aperture in the processing of the received wavefront and does not waste information in centroiding. Furthermore, as the computer power increases every year, the potential for real time implementation will eventually be realized.

Nonetheless, phase retrieval is a non-linear problem but utilizes all the available information as opposed to the Shack-Hartmann sensor which simplifies the problem of phase retrieval to one of solving a system of linear equations. A consequence of dealing with a non-linear problem is there is no guarantee in uniqueness of the solution. The normal approach of solving non-linear equations by optimizing a cost function frequently results in sub-optimal solutions, which is one reason that phase retrieval has not been successfully used in wavefront estimation. This is mainly because quite different phase distributions can result

in very similar distributions in the image plane and when noise is present these distributions become indistinguishable.

Another problem with the use of phase retrieval for wavefront estimation is the existence of an ambiguity in the recovered data. Specifically, if the aperture is symmetric with a phase distribution of $\phi(u,v)$ then $-\phi(-u,-v)$ results in the same observed data. The analog equivalent of this problem occurs in the curvature sensor [55], and in essence means the sign of the even component of the phase cannot be deduced from $d(x,y)$. However, many of the problems in phase retrieval can be overcome if the prior distribution of the phase statistics is known. This prior distribution, by contrast, has been constantly used by the Shack-Hartmann wavefront sensor since there are only a finite number of measurements from the sensor but there is a set of an infinite number of Zernike polynomials to be expanded. The prior information is usually in the form of covariance matrix of the coefficients of the Zernike polynomials.

The motivation for the study described in this chapter is to show how prior knowledge of the statistics of the phase aberrations induced by turbulence can be used to help form an estimate of the phase aberrations introduced by the turbulence. This chapter is organized as follows.

The existing phase retrieval methods are reviewed in section 5.1. These techniques form the basis of the new technique but do not make use of prior statistics. A new technique described in section 5.2 utilizes the Bayesian rule to incorporate the prior probability into the likelihood function being maximized. The procedure is, however, computationally expensive. Hence, for a comparison, an empirical smoothness constraint, which lacks the theoretical basis but is considerably simpler to compute is also discussed in this section. The addition of terms related to the likelihood of a particular phase estimate means that the solutions obtained are fundamentally different from those of projecting onto constraints [71]. This difference is more significant for noisy data when greater weight is placed on the prior statistics of the phase.

Section 5.3 presents the simulation approach used to validate the derived theory. This approach differs from those presented in section 3.2.3 in that it generates the phase directly from the statistics of the turbulence. A square phase screen of 16×16 is chosen in this simulation. This has the disadvantage that the maximum severity of turbulence that can

be simulated adequately is equal to $D/r_0 = 4$. It is, however, also necessary to restrict the image size since a large number of trials were required to obtain statistically significant results.

Numerical results are then presented and discussed in section 5.4. Local maxima still seem to be the major problem in phase retrieval, and they are associated with ambiguities in the reconstructions. Methods for overcoming the local maxima are given in section 5.5. A phase retrieval method using a lenslet array in the manner that the Shack-Hartmann sensor works is introduced in section 5.6. Finally, the chapter concludes with a summary in section 5.7.

5.1 Review of Existing Iterative Phase Retrieval Algorithms

The problem of relating the phase and magnitude of the Fourier transform has received much attention over the years. The foundation for many of the phase retrieval algorithms is the general iterative loop shown in Fig. 5.1 [67], the basic idea for which was originally proposed by Gerchberg and Saxton [6]. The loop basically alternates between image and Fourier space, via Fourier transformation, applying constraints in each domain to force the image estimate towards an image that accurately represents a solution to the particular phase retrieval problem. Iterations are continued until a Fourier pair is found that satisfies the constraints in both domains or until a predetermined number of iterations is completed.

The constraints, usually applied in Fourier space, are the known Fourier magnitude and any partial information known about the phase in the actual image. The most common constraint used in image space is the support constraint. This specifies that the image estimate must be compact (section 2.2). The extent is usually estimated from the auto-correlation [67], and the positivity constraint, that specifies that the image estimate must be non-negative and real [5].

5.1.1 Gerchberg-Saxton algorithm

In this sub-section the Gerchberg-Saxton algorithm is described, since it serves as a useful introduction for the important algorithms described in the following sub-section. The algorithm uses the general iterative loop outlined in Fig. 5.1 and imposes the following image

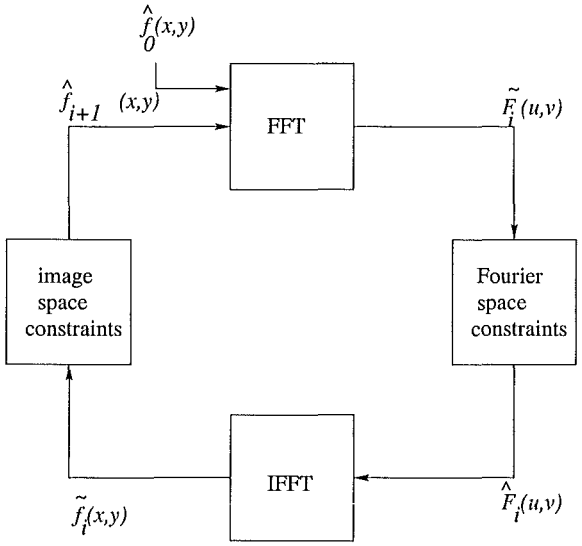


Figure 5.1: *The general iterative phase retrieval loop.*

and Fourier domain constraints in an attempt to recover the unknown phase. In Fourier space the magnitude of the estimated spectrum is constrained in a manner that makes it equivalent to the measured Fourier magnitude, and in image space prior knowledge of the magnitude of $f(x,y)$ is used to constrain the magnitude of the image estimate [6,67]. Thus, one iteration of the Gerchberg-Saxton algorithm can be described by the following four steps:

$$\begin{aligned} \tilde{F}_j(u,v) &= \mathcal{F}\{\hat{f}(x,y)\} \\ \hat{F}_j(u,v) &= |F(u,v)|\exp[i\mathrm{Ph}(\tilde{F}_j(u,v))] \\ \tilde{f}_j(x,y) &= \mathcal{F}^{-1}\{\hat{F}_j(u,v)\} \\ \hat{f}_{i+1}(x,y) &= |f(x,y)|\exp[i\mathrm{Ph}(\tilde{f}_j(x,y))] \end{aligned} \tag{5.1}$$

where a subscript j represents integer indices.

The iteration described by the above set of equations cannot provide estimates that diverge from the actual solution, but this does not guarantee convergence since the algorithm can lock onto a particular estimate which changes very slowly or not at all with each iteration as first shown by Gerchberg and Saxton. The algorithm is then said to have stagnated.

Variants of the Gerchberg-Saxton algorithm, designed to improve its convergence properties, have been reported by several authors [7,67]. A particularly useful set of algorithms, based on the Gerchberg-Saxton algorithm, but with different image space constraints are now

described.

5.1.2 Fienup's algorithms

This sub-section describes iterative phase retrieval algorithms developed by Fienup [67] to solve the phase retrieval problem in the situation where no prior information about $|f(x,y)|$ is available but the image space constraints specified in the introduction of this section are known. The algorithms of Fienup replace the last equation in the set of equations defining the Gerchberg-Saxton algorithm (Eq. (5.1)) by an equation that imposes the image space constraints.

There are two main techniques for imposing the image constraints. The first is to form the next input to the iterative loop by

$$\hat{f}_{i+1}(x,y) = \begin{cases} \tilde{f}_j(x,y) & \text{if not violating any known constraints} \\ 0 & \text{if violating any known constraints,} \end{cases} \tag{5.2}$$

and leads to the error reduction (ER) algorithm. The alternative is to use

$$\hat{f}_{i+1}(x,y) = \begin{cases} \tilde{f}_j(x,y) & \text{if not violating any known constraints} \\ \hat{f}_j(x,y) - \gamma \tilde{f}_j(x,y) & \text{if violating any known constraints,} \end{cases} \tag{5.3}$$

which leads to the hybrid input-output (HIO) algorithm. The parameter γ is a real constant, referred to as the feedback parameter, which is typically set to a value between 0 and 1 [5].

The name "error reduction" comes from the fact that the error in the image estimate must decrease or remain constant with each iteration. Thus, the ER algorithm can stagnate but not diverge. Typically, the error initially decreases quite quickly, but the rate of convergence is very slow [67]. Furthermore, Fienup [67] has shown that the ER algorithm is in fact a steepest descent method, and can be viewed as projecting the direction of steepest descent onto a set of constraints. However, the fact that the ER algorithm uses magnitude constraints, which do not form convex sets, precludes the algorithm from being classified as a method of Projection Onto Convex Sets (POCS) [7, 67, 71]. This is unfortunate as it precludes guaranteed convergence to a unique global maximum.

The HIO was developed to speed up the rate of convergence of the ER algorithm. In contrast, with the HIO there is no guarantee that the error decreases with each iteration and it tends to fluctuate as the algorithm converges to a solution. This instability can

be a problem when the Fourier magnitude is very noisy, since the HIO can diverge after coming close to the true image [5]. HIO is particularly effective for positive images, but can on occasion stagnate. Three causes of stagnation have been identified by Fienup and Wackerman [7], who have developed methods for avoiding such stagnation in the context of positive images. On the other hand, the HIO is reported to be the most effective of all the iterative phase retrieval algorithms as it usually converges to a solution much faster than the other techniques [5, 7, 67].

5.2 Proposed Method by use of a Statistical Approach

This section introduces a new approach to the solution of two-dimensional phase retrieval problems. The technique is based on the concept of how the statistics of Kolmogorov turbulence can be used to regularize the phase retrieval problem which is generally ill-posed. The statistics are then used to compute the likelihood for a particular phase screen. A similar method has also been used for the wavefront reconstruction using the Shack-Hartmann sensor, which is described in the previous chapter.

The wavefront transmission of the instrument pupil is modeled by a complex function:

$$A(u,v) = M(u,v)\ \exp[i\phi(u,v)]. \tag{5.4}$$

Assume that scintillation effects are small so $M(u,v)$ can be computed using knowledge of the instrument geometry and the number of photons received. The phase aberration $\phi(u,v)$ is a random field assumed to arise from the passage of light through Kolmogorov turbulence (section 3.2). This assumption is not restrictive because the results presented here can easily be extended to other turbulence models or to the statistics measured from turbulence [59].

Assuming incoherent, noise free imaging of a point source, the available measurements in the image plane are equal to (section 3.1)

$$\mu(x,y) = |a(x,y)|^2 \tag{5.5}$$

where $a(x,y)$ and $A(u,v)$ form a Fourier transform pair

$$\begin{aligned} A(u,v) &= \mathcal{F}[a(x,y)] \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} a(x,y)\exp[-i2\pi(ux+vy)]\mathrm{d}x\mathrm{d}y. \end{aligned} \tag{5.6}$$

Assume that the noise on the image results from both a finite number of photons being received by the instrument and from spontaneous independent events from the detector at each pixel location. The resulting noise is the sum of two independent Poisson processes. The detected image of the point source $d(x,y)$ is thus given by

$$d(x,y) = \mu(x,y) + \zeta(x,y) + n(x,y), \tag{5.7}$$

where $\mu(x,y)$ is the mean of the received photons, $\zeta(x,y)$ is the mean of the independent events produced by the detector (assumed to be known) and $n(x,y)$ is the zero mean random noise component formed from combining both the photon and detector noise.

The comparison of an estimated phase $\hat{\phi}(u,v)$ to the true phase $\phi(u,v)$ is made more complicated by the existence of a number of ambiguities. There are three main changes that can be made to $\hat{\phi}(u,v)$ that have no effect on the estimated mean $\hat{\mu}(x,y)$;

1. Addition of a constant to $\hat{\phi}(u,v)$
2. Addition of a multiple of 2π to any point in $\hat{\phi}(u,v)$
3. Replacing $\hat{\phi}(u,v)$ by $-\hat{\phi}(-u,-v)$ [provided $M(u,v)$ is symmetric].

Since these changes can not be recovered from the observed data $d(x,y)$, it is important to use a comparison between the true and estimated phase which is independent of these ambiguities.

The simplest method for removing the effects of the first two ambiguities in a comparison is the use of the Strehl ratio as defined in Eq. (3.18). The Strehl ratio corresponds well to the mean squared error (mse) in the phase after the first two ambiguities are eliminated. The last ambiguity requires $\hat{\phi}(u,v)$ or $-\hat{\phi}(-u,-v)$ to produce the highest Strehl ratio. This last ambiguity is a common problem with phase retrieval and can be addressed with the introduction of phase diversity as in the curvature sensor [60] although it is common practice in the comparison of phase retrieval algorithms to select either $\hat{\phi}(u,v)$ or $-\hat{\phi}(-u,-v)$ on the basis of which one produces the highest Strehl ratio.

A simple statistical statement of the phase retrieval problem is to find the estimate $\hat{\phi}(u,v)$ that is most likely given the observed data $d(x,y)$. In the formal mathematical formulation,

it can be expressed as

$$\max_{\hat{\phi}} Pr\{\hat{\phi}(u,v)|d(x,y)\}. \tag{5.8}$$

Note that this technique is also used for wavefront estimation using the Shack-Hartmann sensor to compute the optimal reconstructor described in the previous chapter (cf. Eq. (4.43)).

Using Bayes theorem described in section 2.3.4 (see also Eq. (4.44)),

$$Pr\{\hat{\phi}(u,v)|d(x,y)\} = \frac{Pr\{d(x,y)|\hat{\phi}(u,v)\}Pr\{\hat{\phi}(u,v)\}}{Pr\{d(x,y)\}} \tag{5.9}$$

and taking the logarithm yields

$$\log[Pr\{\hat{\phi}(u,v)|d(x,y)\}] = \log[Pr\{d(x,y)|\hat{\phi}(u,v)\}] + \log[Pr\{\hat{\phi}(u,v)\}] - \log[Pr\{d(x,y)\}]. \tag{5.10}$$

The final term on the right hand side can be discarded from further consideration since it is not a function of the phase estimate $\hat{\phi}(u,v)$. For notation convenience $\log[Pr\{d(x,y)|\hat{\phi}(u,v)\}]$ in Eq. (5.10) is referred to as LL.

Two basic algorithms are considered here to describe this approach. The first is a maximum likelihood (ML) algorithm [61] formed by maximizing Eq. (5.10) with $\log[Pr\{\hat{\phi}(u,v)\}]$ assumed to be constant. The second is a maximum a posteriori (MAP) estimate which can be formed when statistical information on $\phi(u,v)$ is available. There are a wide variety of possibilities for maximizing the resulting log-likelihoods, but one of the simplest is to use a standard optimization package such as the NAG libraries [62]. These require the evaluation of the gradient of the log-likelihood.

Considering the ML case first, the gradient is obtained by differentiating Eq. (5.10),

$$\frac{\partial LL}{\partial \hat{\phi}(u,v)} = \frac{\partial LL}{\partial \hat{a}(x,y)} \frac{\partial \hat{a}(x,y)}{\partial \hat{A}(u,v)} \frac{\partial \hat{A}(u,v)}{\partial \hat{\phi}(u,v)}. \tag{5.11}$$

The next stage in this approach is to incorporate the statistical model for the noise on the data $d(x,y)$. Assume the Poisson noise model discussed earlier (cf. Eq. (4.58)),

$$Pr\{d(x,y)|\hat{\phi}(u,v)\} = \prod_{x,y} \frac{(\hat{\mu}(x,y) + \zeta(x,y))^{d(x,y)} \exp[-\hat{\mu}(x,y) - \zeta(x,y)]}{d(x,y)!}, \tag{5.12}$$

where $\hat{\mu}(x,y)$ is computed from Eqs. (5.4),(5.5) and (5.6). Taking the log of this equation and discarding terms which are not functions of $\hat{\phi}(u,v)$ yields

$$LL = \sum_{x,y} d(x,y) \log[\hat{\mu}(x,y) + \zeta(x,y)] - \hat{\mu}(x,y) - \zeta(x,y). \tag{5.13}$$

The derivative calculation follows that in Thiebaut and Conan [63],

$$\frac{\partial LL}{\partial \hat{\phi}(u,v)} = \mathrm{Im}\left[\mathcal{F}\left(\frac{d(x,y) - \hat{\mu}(x,y) - \zeta(x,y)}{\hat{\mu}(x,y) + \zeta(x,y)}\, a(x,y)\right) A^*(u,v)\right]. \tag{5.14}$$

In addition, the MAP estimate requires the evaluation of the probability of $\hat{\phi}(u,v)$, and its derivative with respect to changes in $\hat{\phi}(u,v)$. The computation of the former presented here follows the work of Wallner [9] and requires the definition of an aperture weighting function in the two-dimensional aperture plane, W_A. It has the property,

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_A(u,v) \mathrm{d}u\mathrm{d}v = 1, \tag{5.15}$$

and is constant within the region A. The phase distortion in this aperture is assumed to have the Kolmogorov structure function typical of atmospheric turbulence as adapted from Eq. (3.34)

$$\mathcal{D}(u_1,v_1,u_2,v_2) = 6.88\left(\frac{|(u_1,v_1) - (u_2,v_2)|}{r_0}\right)^{5/3}. \tag{5.16}$$

From Eq. (5.16) it can be seen that if r_0 is small the phase in the aperture decorrelates over a short distance.

The statistics defined in Eq. (5.16) indicate that the phase distortion in the telescope aperture is fractal in nature. A consequence of this is that the covariance of the phases in the aperture can only be defined for (cf. Eq. (3.47))

$$\varphi(u,v) = \phi(u,v) - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi(u,v) W_A(u,v) \mathrm{d}u\mathrm{d}v. \tag{5.17}$$

This is not a serious problem in practice as the average value of $\hat{\phi}(x,y)$ does not affect the observed data $d(x,y)$ as described earlier in section 3.2.3.

The second order statistics of the phase can then be encapsulated in the covariance matrix as (section 2.3.5)

$$\langle \varphi(u_1,v_1)\varphi(u_2,v_2)\rangle = C(u_1,v_1,u_2,v_2). \tag{5.18}$$

This can be computed by [9]:

$$C(u_1,v_1,u_2,v_2) = -\frac{1}{2}\mathcal{D}(u_1,v_1,u_2,v_2) + g(u_1,v_1) + g(u_2,v_2) - b, \tag{5.19}$$

where

$$g(u,v) = \frac{1}{2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_A(u',v')\mathcal{D}(u,v,u',v')\mathrm{d}u'\mathrm{d}v' \tag{5.20}$$

and

$$b = \frac{1}{2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_A(u',v')W_A(u'',v'')\mathcal{D}(u',v',u'',v'')\mathrm{d}u'\mathrm{d}v'\mathrm{d}u''\mathrm{d}v''. \tag{5.21}$$

For the computation of $\log[Pr\{\hat{\phi}(u,v)\}]$ it is convenient to first rearrange the two-dimensional image as a single vector by either rows or columns [66]. $\hat{\mathbf{\Phi}}$ is used to represent the reordered phase vector and $\mathbf{C}$ to represent the matrix of covariances between the elements of $\hat{\mathbf{\Phi}}$ computed using Eq. (5.20). The probability $\log[Pr\{\hat{\phi}(u,v)\}]$ can then be expressed as [3]

$$\log[Pr\{\hat{\phi}(u,v)\}] = \frac{1}{2}\hat{\mathbf{\Phi}}^T\mathbf{C}^{-1}\hat{\mathbf{\Phi}}. \tag{5.22}$$

The derivative with respect to the phase is equal to

$$\frac{\partial \log[Pr\{\hat{\phi}(u,v)\}]}{\partial\hat{\phi}(u,v)} = \mathcal{R}^{-1}[\hat{\mathbf{\Phi}}^T\mathbf{C}^{-1}], \tag{5.23}$$

where $\mathcal{R}^{-1}[]$ indicates that the one dimensional reordering described above is to be reversed to produce a two-dimensional quantity.

It is readily apparent that Eq. (5.22) is computationally expensive not only because of the storage requirements of $N^2 \times N^2$ matrix $\mathbf{C}$, where $N \times N$ is the dimension of a phase screen, but also because of the extremely long times required to perform the matrix inversion. On the other hand, once the $\mathbf{C}^{-1}$ matrix has been calculated and stored, computation of Eq. (5.22) requires only a matrix multiplication. It is, however, reasonable to consider whether a computationally less demanding smoothness constraint may produce the same effect as the MAP estimate. Therefore, the MAP estimate is compared with a penalized ML estimate (ML+) which is common regularization in practice [23]. By using this method, the log-likelihood to be maximized becomes:

$$LL_p = LL + \gamma\Gamma(u,v) \tag{5.24}$$

where $\Gamma(u,v)$ is a penalty term that is a function of the estimated phase $\hat{\phi}(u,v)$. The non-negative scale factor γ determines how strongly the penalty is enforced. It is usual to select a penalty that both ensures smoothness of the estimated phase and simplifies the maximization of LL_p. It is reasonable to use the curvature of the phase estimated by using

a Laplacian [23], since this is widely used in function estimation problems. The Laplacian of the phase is defined by

$$\Gamma(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nabla^2\hat{\phi}(u,v)\mathrm{d}u\mathrm{d}v, \tag{5.25}$$

where ∇ is the gradient operator, and it is implemented digitally by convolving the phase with the convolution kernel

$$\ell = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}. \tag{5.26}$$

Thus, Eq. (5.24) now becomes

$$LL_p = LL + \gamma\sum_{u,v}\{\hat{\phi}(u,v) \odot \ell\}^2. \tag{5.27}$$

The derivative with respect to the estimated phase is simply

$$\frac{\partial LL_p}{\partial\hat{\phi}(u,v)} = \frac{\partial LL}{\partial\hat{\phi}(u,v)} + 2\gamma\{\hat{\phi}(u,v) \odot \ell\} * \ell. \tag{5.28}$$

It is important to note that other penalty functions can be used and that the choice of $\Gamma(u,v)$ in Eq. (5.24) is dictated by convenience, and is in no way claimed to be optimal.

It is important to emphasize that maximization of the log-likelihood by conventional methods does not guarantee solution of the phase retrieval problem because of the existence of local maxima in addition to the global maximum of LL. It is worth noting that the prior terms defined in Eqs. (5.22) and (5.25) do have unique global maxima, and their addition to LL is likely to reduce the severity of local maxima.

5.3 Simulation Approach

The phase distortion can be simulated using a number of alternative approaches (section 3.2.3) but in this case the phase is generated directly by the use of eigensystem of $\mathbf{C}$ [70]

$$\phi(u,v) = \sum_j \sqrt{c_j} r_j v_j(u,v) \tag{5.29}$$

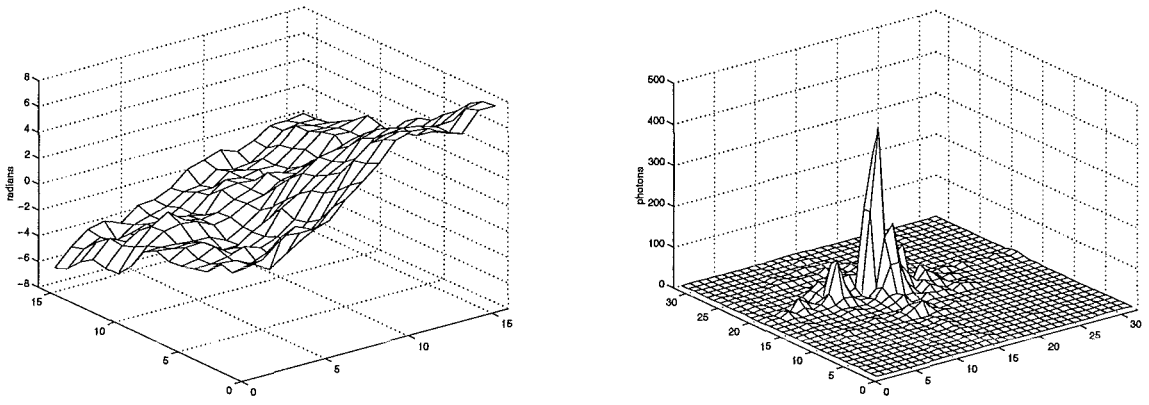
where $v_j(u,v)$ and c_j are the j^{th} eigenvector and eigenvalue of the covariance matrix $\mathbf{C}$, respectively. The vector r_j is comprised of Gaussian random numbers of zero mean and unit variance [3]. It is worth noting that essentially this approach is the same as that of Roddier's method described in section 3.2.3 but uses a different set of basis functions (cf.

Eq. (3.44)). However, this direct technique has the advantage of producing an exact phase screen since there is no truncation in the basis functions as opposed to the Roddier method in which it is impossible to use an infinite number of Zernike basis functions.

The total phase variance over a square aperture of dimension D (see Eq. (4.33) for a circular aperture) is given by [38]

$$1.31(D/r_0)^{5/3}, \tag{5.30}$$

a substantial portion of which can be attributed to an overall slope in the phase. This overall slope is clearly evident in the phase screen shown in Fig. 5.2a, with the corresponding simulated data for $D/r_0 = 4$ and 5000 photons shown in Fig. 5.2b. It is, however,



(a) (b)

Figure 5.2: *(a) A simulated phase screen $D/r_0 = 4$ (b) the corresponding noise-corrupted image for 5000 photons.*

inappropriate to rely on the results obtained from a single speckle image since there is a substantial statistical variation present in both the phase and the Poisson noise. The results in this thesis are presented for an ensemble of speckle images in order to produce conclusive differences between the methods.

The non-convexity of the phase problem usually results in Eq. (5.9) having multiple maxima. A consequence of this is the performance also depends significantly on the choice of starting point. This proves to be even more critical in the case of phase retrieval using conventional maximization strategies since, as is shown in the next section, a poor choice of starting point can actually cause the addition of prior information to degrade the reconstruction from a phase retrieval algorithm. The initial trials were started at three different points:

zero phase, a linear approximation to the phase computed from the centroid of $d(x, y)$, and the true phase $\phi(u, v)$. The Strehl ratios obtained from these starting points are referred to as S_z, S_l and S_t, respectively. The log-likelihoods corresponding to these Strehl ratios are denoted by LL_z, LL_l and LL_t.

The zero phase starting estimate is one that has often been used in phase retrieval and often results in a failure to converge to the global maximum of the likelihood. The linear phase starting estimate is a simple improvement which can be computed from the centroid of the data. This has often been employed in simple adaptive optics systems [42]. The true phase can not be used as a starting point in practice, but provides a useful perspective into the feasibility of phase retrieval since it gives a measure of the best result that can possibly be obtained. It should be noted that in practice even this starting estimate may result in an unsatisfactory approximation to the true phase.

Clearly, if a phase retrieval algorithm diverges significantly under the influence of noise when started from the true phase, the algorithm is impractical. Another point is that a local maximum found near the true phase may have a lower likelihood than that found from the other starting estimates. This presents a major problem because it implies that even with an optimization strategy capable of overcoming local maxima, the task is hopeless as the true phase is not close to the global maximum of the likelihood function.

Figure 5.3 shows three reconstructions from a starting point of a linear phase approximation using ML, ML+ and MAP, respectively. For ML+ the parameter γ was varied between 1 and 50, with the optimum value of $\gamma = 5$ determined experimentally. Unfortunately, there is no definitive way to find the optimum weight, a significant problem with this regularization technique. It is noticeable that the ML+ is ineffective at enforcing the smoothness constraint at the cost of the edges. It is also obvious that the MAP produces an estimate which is close to the true solution and the ML+ gives a smoother version of the ML estimate.

5.4 Numerical Results

The initial results, averaged over 1000 speckle images, are shown in table 5.1 with the corresponding standard deviation of these results shown in table 5.2. All algorithms were terminated at 200 iterations.

Photons	Starting phase								
	Zero			Linear			True		
	ML	ML+	MAP	ML	ML+	MAP	ML	ML+	MAP
50	0.1804	0.0619	0.1229	0.1889	0.1571	0.2417	0.3074	0.5735	0.5004
500	0.2266	0.1058	0.1888	0.3209	0.2223	0.3692	0.7472	0.7452	0.8357
5000	0.3087	0.1768	0.2136	0.3428	0.3274	0.3403	0.9586	0.9289	0.9646

Table 5.1: *Mean Strehl ratios obtained for phase retrieval for speckles formed with $D/r_0 = 4$. Results averaged over 1000 different noise-corrupted speckle images.*

Photons	Starting phase								
	Zero			Linear			True		
	ML	ML+	MAP	ML	ML+	MAP	ML	ML+	MAP
50	0.0036	0.0039	0.0032	0.0036	0.0076	0.0051	0.0035	0.0061	0.0043
500	0.0057	0.0058	0.0041	0.0063	0.0171	0.0078	0.0015	0.0045	0.0011
5000	0.0072	0.0100	0.0049	0.0077	0.0290	0.0077	0.0002	0.0012	0.0002

Table 5.2: *Standard deviations of the mean Strehl ratio values reported in table 5.1.*

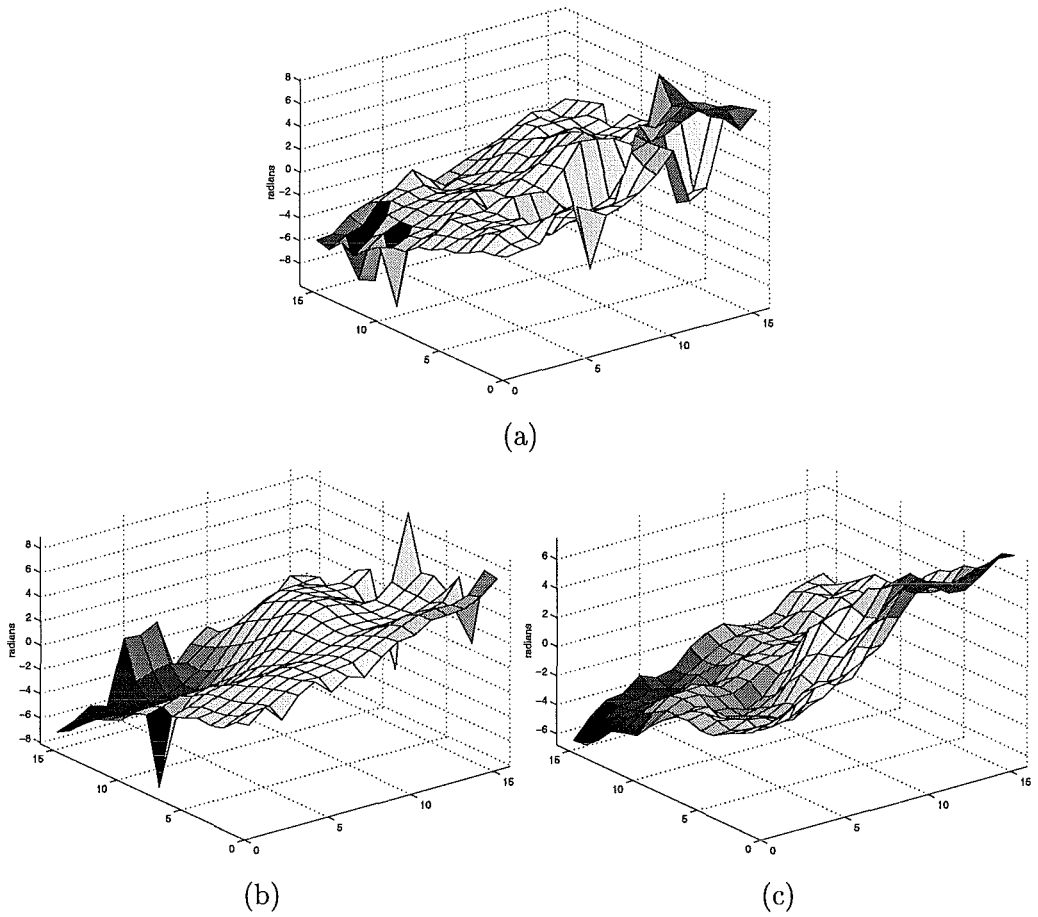


Figure 5.3: *Phase estimates produced by (a) the conventional ML method (b) the ML plus a penalty function and (c) the MAP estimate. In all cases the linear phase starting point is used, $D/r_0 = 4$ and there are 5000 photons.*

It is readily apparent that ML+ algorithm produces the highest standard deviations and therefore has the least consistent reconstruction performance. ML and MAP algorithms, however, are comparable from this perspective. A result of note is that the Strehl ratio achieved by the ML, ML+ and MAP methods when started from either zero or a linear phase is significantly lower than that achieved from starting from the true phase. This indicates that convergence is occurring at a local maximum in the log-likelihood.

It can be seen that ML+ achieves an improvement over ML only for the true phase starting estimate and 50 photons. One of the reasons that ML+ does not produce the same quality as the MAP estimate is because the penalty function used is only a local operator and unable to enforce any long distance correlations.

The difference in Strehl ratio achieved from the different starting points is attributable to local maxima, although the ML and MAP suffer from different types of local maxima. The presence of local maxima is evident in Fig. 5.4 which shows $S_t - S_l$, the difference in Strehl ratio obtained from the true and linear phase starting points for the algorithms. In very few

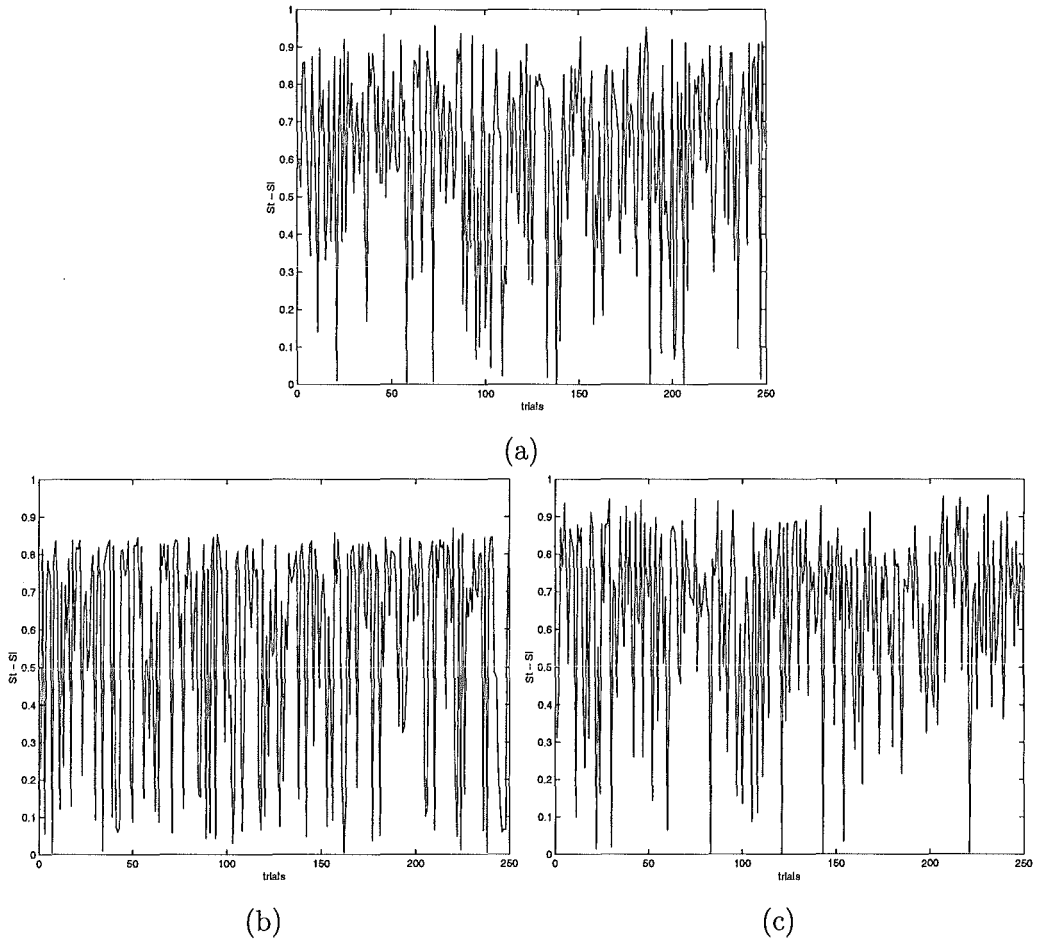


Figure 5.4: *$S_t - S_l$, the difference between the the Strehl ratio obtained from a linear phase starting estimate and S_t obtained from starting at the true phase for (a) ML (b) ML+ and (c) MAP. $D/r_0 = 4$ and 5000 photons. Note that only in a few cases are the differences close to zero indicating convergence to the same local maximum in the likelihood function.*

of the trials was the difference in achieved Strehl ratio sufficiently close to zero to indicate that convergence to the same maximum in the likelihood function had occurred. In all other cases the Strehl ratio obtained from a linear starting phase is significantly worse than what could potentially be obtained if the global, and not a local, maximum in the likelihood was

located.

5.4.1 Local maxima in ML phase estimates

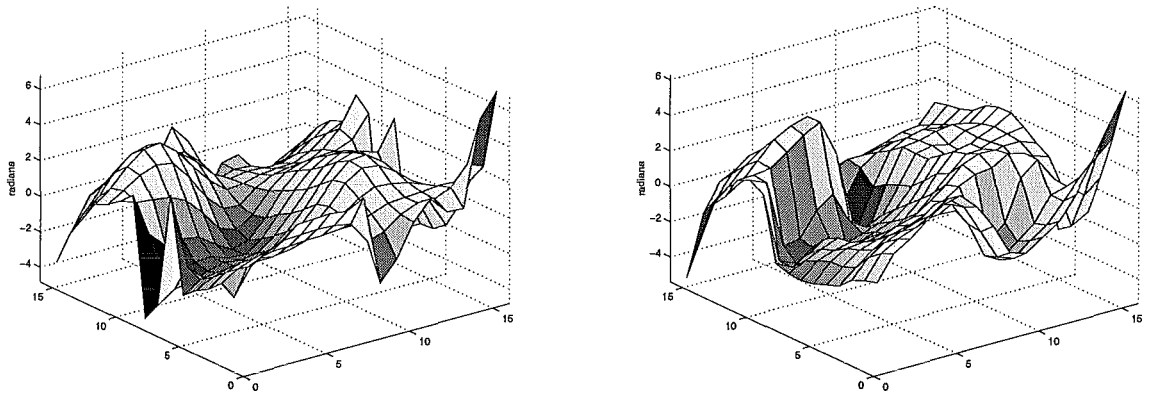
There are two main sources of local maxima in the ML phase retrieval problem. The first is the non-convexity of the Fourier magnitude constraint, which produces these local maxima even in the absence of noise. The second is points where the Fourier magnitude is zero or dominated by noise. The latter causes an ill-posed problem since any phase attributed to these points produces a reconstruction consistent to within the noise on the observed data $d(x, y)$. The problem of many phase screens being capable of producing the data to within the bounds posed by the expected noise requires extra information to ensure a reliable solution.

The addition of a prior distribution in the MAP technique stabilizes the solution and enables a choice to be made between two possible phase distributions which are consistent with the observed data. The decision is made statistically so it is unlikely that the distribution further from the true solution is chosen.

It does appear, however, that the addition of prior information also reduces the severity of the local maxima due to the non-convexity. This is because the prior term $\hat{\mathbf{\Phi}}^T \mathbf{C}^{-1} \hat{\mathbf{\Phi}}$ in Eq. (5.22) has a unique well defined maximum. As the noise on $d(x, y)$ increases, the relative weight on the prior also increases, simplifying the maximization of Eq. (5.9), and consequently reducing the probability of being trapped in a local maximum. At very low noise levels the relative weight on the prior term is small and consequently the difference between ML and MAP algorithms is negligible.

5.4.2 Regularization induced local maxima

The ML+ and MAP approaches do, however, introduce another source of local maxima in addition to those of the ML approach. This is apparent in Fig. 5.5 which shows both ML+ and MAP estimates formed at $D/r_0 = 4$ with 5000 photons from a zero phase starting point. The existence of steep changes in Fig. 5.5a and Fig. 5.5b corresponding to 2π discontinuities in the reconstructed phase is readily apparent. The local maxima associated with these changes cause the ML+ and MAP algorithms to perform worse than the ML algorithm when the starting point is significantly removed from the true solution. This is evident in



(a) (b)

Figure 5.5: *Illustration of the wrapping effects when a zero phase starting point is used, (a) the ML+ estimate (b) the MAP estimate.* $D/r_0 = 4$ *and 5000 photons.*

the first column of table 5.1, which shows the performance from a zero phase starting point. In this case the performances of the ML+ and MAP algorithms are significantly worse than the performance of the ML algorithm.

The cause of this performance difference is that while the relationship over short distances in $\hat{\phi}(u,v)$ is substantially correct, the long distances which contribute to the high frequencies of $\hat{\mu}(x,y)$ are substantially in error due to a 2π wrapping. The difficulty is that although the likelihood would be increased by removing the 2π steps, the MAP algorithm relies on gradient search techniques making incremental changes in $\hat{\phi}(u,v)$. Unfortunately, any small change in the phases actually results in a decrease in the likelihood, and so a gradient based technique remains trapped in a local maximum. At lower photon counts, the effect of the prior is to smooth the region of the discontinuity, but the presence of a local maximum in the likelihood function remains.

5.4.3 Correlation between the log-likelihood and Strehl ratio

An important feature of any statistically based phase retrieval algorithm is how well the maximum in the likelihood function corresponds to the error in the phase reconstruction [5]. This leads to a further important difference between the ML, ML+ and MAP approaches shown in Fig. 5.6. The abscissa of the graphs show $LL_t - LL_z$, the difference of the likelihoods obtained when starting at the true phase and a zero phase starting estimate for 50 photons. The ordinate shows $S_t - S_z$, the difference in the Strehl ratio obtained

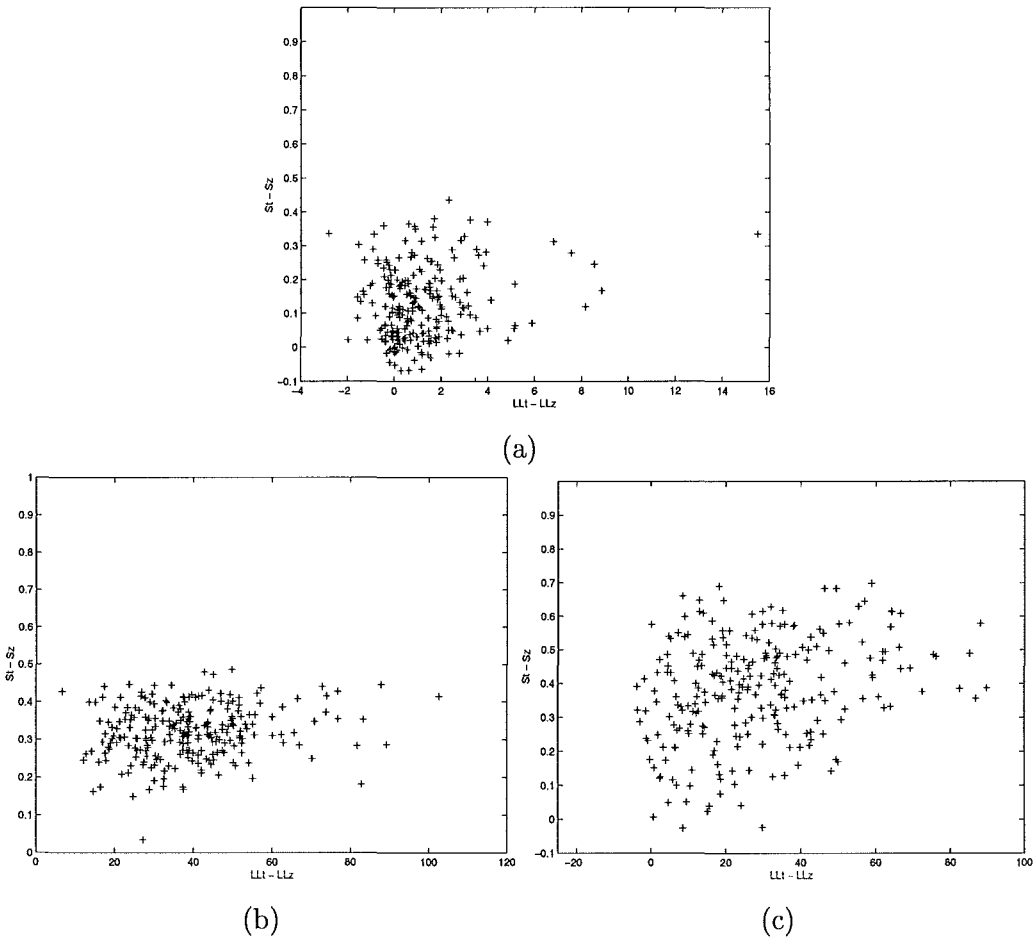


Figure 5.6: *$S_t - S_z$ plotted as a function of $LL_t - LL_z$. (a) ML (b) ML+ (c) MAP. Note in particular the difference in scales of the two graphs. Also note the large number of cases in the ML algorithm when either $S_t - S_z$ or $LL_t - LL_z$ is negative, indicating the global maximum of the likelihood is displaced from the true phase.*

from starting at the true phase and starting at zero phase. There are two major problems of note. First, there is a poor correlation between the difference in Strehl ratio and the difference in likelihoods obtained; and this is significantly worse for the unsmoothed ML estimates. Second, there is a large number of ML trials where the likelihood of the phase recovered from the zero phase starting point is higher than that recovered from the true phase starting point, despite having a lower Strehl ratio. This poses a significant problem since it means the global maximum does not correspond to a good approximation of the true phase. This problem is exacerbated by increasing D/r_0 and fewer photons. The MAP algorithm, by contrast, has a comparatively smaller number of reconstructions where the true phase starting point produces a lower likelihood or a lower Strehl ratio than the zero

phase starting point. In this respect the ML+ algorithm was slightly superior to MAP because all reconstructions starting from the true phase resulted in a higher Strehl ratio than from the zero phase starting point.

The true phase is, of course, an unrealistic starting point but it can be argued that in many cases a good estimate of the phase may be available. As an example, when monitoring time evolving turbulence the previous phase estimate would be a good starting point, provided the spacing between temporal samples is reasonably short. In this case phase retrieval would be useful provided there was a region around the true phase where convergence would occur.

5.4.4 Region of convergence for phase retrieval

The region of convergence for phase retrieval was tested by starting the ML, ML+ and MAP algorithms from a point perturbed from the true phase. 1000 different speckle images were generated corresponding to $D/r_0 = 4$. The starting point used was the true phase perturbed by adding another phase screen with a random D/r_0 between 0 and 4.

Figure 5.7 shows the attained Strehl ratio as a function of the mean squared of the added phase perturbation. It is clear that for low perturbations, both the ML and MAP converge consistently to a solution yielding a high Strehl ratio. This corresponds to a range of 1.5 and 3 for the ML and MAP, respectively, illustrating that the addition of a statistical prior increases the permissible error in the starting point which still yields acceptable convergence.

Whilst the ML+ also yields a region of convergence it is both narrower than that of ML and in addition the Strehl ratio attained in this region is lower. The lower Strehl ratio is not unexpected since the ad-hoc prior can be expected to cause some distortion of the phase estimate.

5.5 Overcoming the Ambiguities

5.5.1 Rotational ambiguity

Although the previous result indicates that phase retrieval, and the MAP algorithm in particular, is applicable when a good starting phase estimate is available, ideally one would prefer to attain a higher convergence rate from the linear phase starting point. Figure 5.8 shows the attained Strehls of 1000 speckle images having $D/r_0 = 1$ when there are 50, 500,

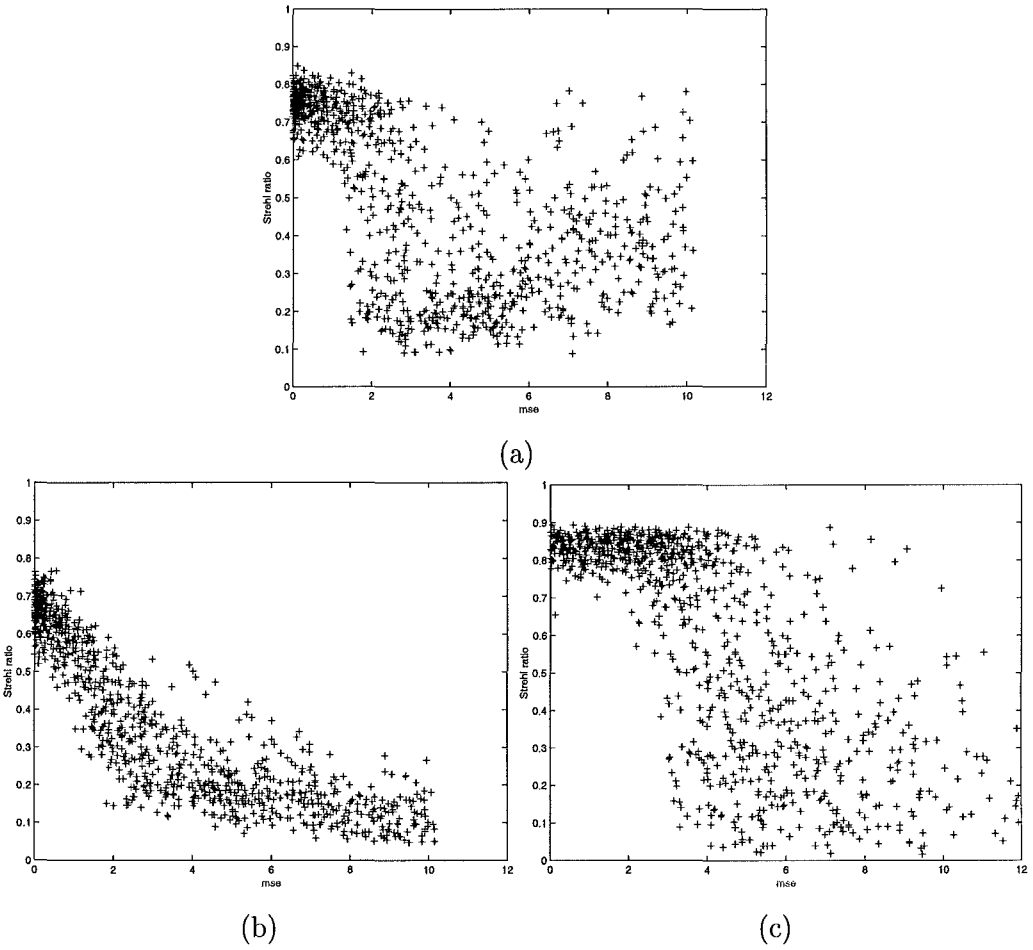


Figure 5.7: *Plot of the attained Strehl ratio as a function of the mse of the starting phase estimate (a) ML, 500 photons, (b) ML+, 500 photons (c) MAP, 500 photons.*

and 5000 photons per image, respectively. It is, thus, clear that, apart from the rotational ambiguity (cf. Fig. 5.8c) for low levels of phase distortion convergence to a high Strehl occurs consistently. It may be noted that the ambiguity is less obvious in Figs. 5.8a and 5.8b. This results from the high weighting on the prior making the recovered phase more planar, and consequently less affected by the 180^0 rotation. This produces the interesting effect that the mean Strehl for 5000 photons (0.90) is actually lower than for 500 photons (0.92), simply because the former suffers more from the rotational ambiguity. One solution to the rotational ambiguity is by use of phase diversity proposed originally by Gonsalves [16] and later generalized by Paxman *et al* [17] for more than two images. One of these images is the conventional focal-plane image that has been degraded by the unknown aberrations. Additional images by the same object are formed by perturbing these unknown aberrations

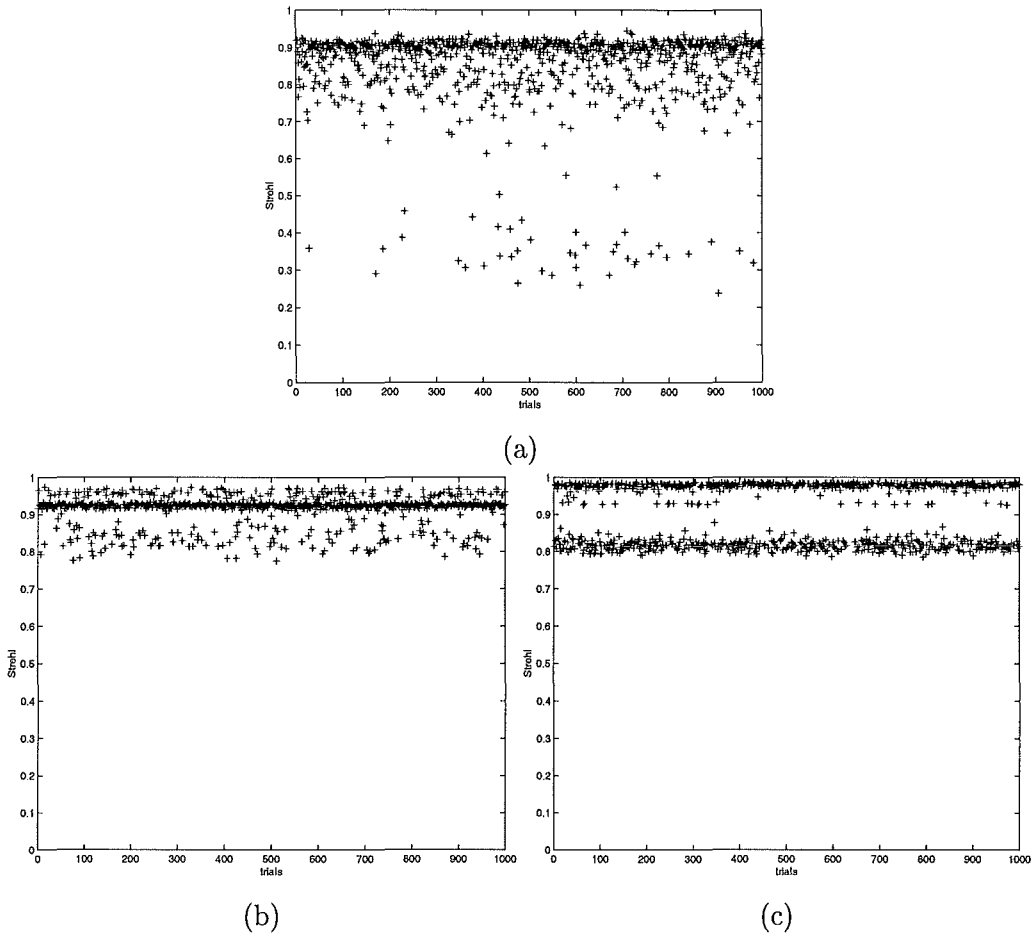


Figure 5.8: *Plot of the attained Strehl ratio by the MAP algorithm as a function of trials for $D/r_0 = 1$ and (a) 50 photons (b) 500 photons (c) 5000 photons.*

in some known fashion. For example, a simple translation of the detector array along the optical axis further degrades the imagery with a known amount of defocus, a point illustrated in Fig. 5.9. Once these estimates are known, the log likelihood function LL_d to be maximized is now the sum of each log likelihood;

$$LL_d = LL + LL_{\text{defocus}}, \tag{5.31}$$

where LL_{defocus} is the log likelihood of the diversity image that is degraded by a known amount of defocus. Simulation studies have shown that this does not significantly affect the rotational ambiguities described here [15].

Another technique to reduce the rotational ambiguity is by subdividing the aperture in the manner of traditional wavefront sensors. This technique is described in section 5.6.

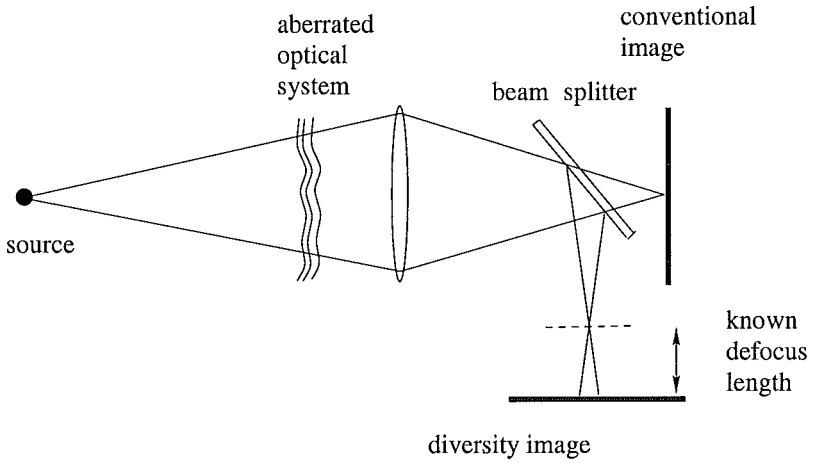


Figure 5.9: *Optical layout of a phase diversity system. The conventional image is degraded by aberrations in the optical system. The diversity image is degraded by the combination of the same aberrations and a known amount of defocus.*

5.5.2 Wrapping ambiguity

As described in section 5.4, ML+ and MAP approaches suffer a significant problem due to steep changes corresponding to 2π discontinuities in the reconstruction. The relationship between the wrapped phase and the unwrapped phase may be stated as

$$\phi_w(x,y) = \phi(x,y) + 2\pi m(x,y), \tag{5.32}$$

where $\phi_w(x,y)$ is the wrapped phase, $\phi(x,y)$ is the unwrapped phase, and $m(x,y)$ is an integer-valued correcting field. Unwrapping is therefore a simple matter of adding or of subtracting 2π offsets at each discontinuity encountered in the phase data. However, unwrapping becomes more difficult when noise is present. A number of solutions addressing this problem have been presented [27–29]. Ghiglia and Romero [28] have proposed an iterative weighted least-squares method for synthetic aperture radar (SAR) images, where regions with a low signal-to-noise ratio (SNR) can be given less weight than consistent region. Friedlander and Francos [27] have used a model based on polynomials to tackle this problem. Pritt and Shipman [29] propose the discrete cosine transform (DCT) as a fast solution of the resulting system of equations. The major drawbacks of these existing methods for phase unwrapping are that they are computationally demanding [27], application dependent [28], or are not optimum in the presence of noise [29].

One approach to this problem would be to try multiple starting points, but the number required rises rapidly with D/r_0, resulting in a dramatic increase in computation. Simulated

Photons	Mean Strehl	σ
50	0.2142	0.0168
500	0.4629	0.0277
5000	0.5969	0.0323

Table 5.3: *Performance in terms of Strehl ratio for MAP phase retrieval with an annealed weighting on the prior. Results should be compared to the MAP estimates from linear phase in tables 5.1 and 5.2.*

annealing [8], which is widely used, is also subject to a heavy computational burden. A possible solution of removing both the rotational and wrapping ambiguities by subdividing the aperture in the manner of traditional wavefront sensors is described in the next section.

In this thesis, a method based on increasing the weight on the prior for the initial iterations is proposed. This effectively constrains the phase to a very smooth solution and makes the production of discontinuities in the phase much less likely. Once convergence of the MAP is achieved the resulting phase can be assumed to provide a better estimate of the true phase. The weighting on the prior is then reduced and a new cycle of the MAP algorithm restarted. This can be viewed as a form of multi-resolution phase recovery since the roughness of the phase is allowed to increase at each cycle. Multi-resolution has been employed in crystallography with some success [68], and the approach proposed here can be viewed as a combination of this method and simulated annealing [8].

Table 5.3 shows the performance improvement obtained using this approach in conjunction with a linear phase starting estimate. The additional weighting of the prior was stepped down in 4 stages of 10, 5, 2 and finally 1. Compared with the MAP estimates starting from linear phase presented in table 5.1 and 5.2 with the exception of 50 photons, it is obvious that the results are much better in Strehl ratio, although having higher standard deviations. At low photon counts placing the prior is already very highly weighted, and little benefit accrues from increasing its weight further. In fact, too high weight on the prior may initially cause the phase estimate to tend to zero.

More importantly, Fig. 5.10 shows 1000 trials at $D/r_0 = 4$ where the performance from a linear starting phase estimate S_l is shown. In order to clarify the display of the results, the trials have had the rotational ambiguity removed and have been sorted in order of ascending Strehl ratio before display. The distinct knee in the curve around trial number 500 corre-

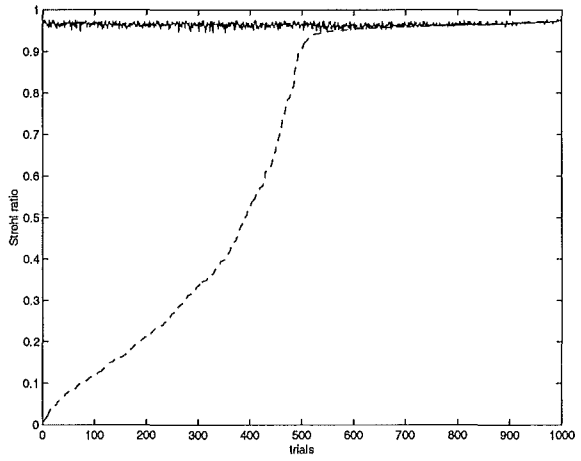


Figure 5.10: *Performance of the MAP algorithm using a variable weighting on the prior probability. The dashed line shows 1000 trials of the variable weighting scheme sorted in ascending order of Strehl ratio obtained. The solid line shows the Strehl ratio obtained from the true phase for comparison.*

sponds to where there is no longer a significant difference between S_l and S_t achieving Strehl values greater than 0.90. In general, convergence to a phase with a wrapping ambiguity is less likely if an evolving prior is incorporated in the maximization. The results show that the use of prior information can significantly improve the phase estimates obtained.

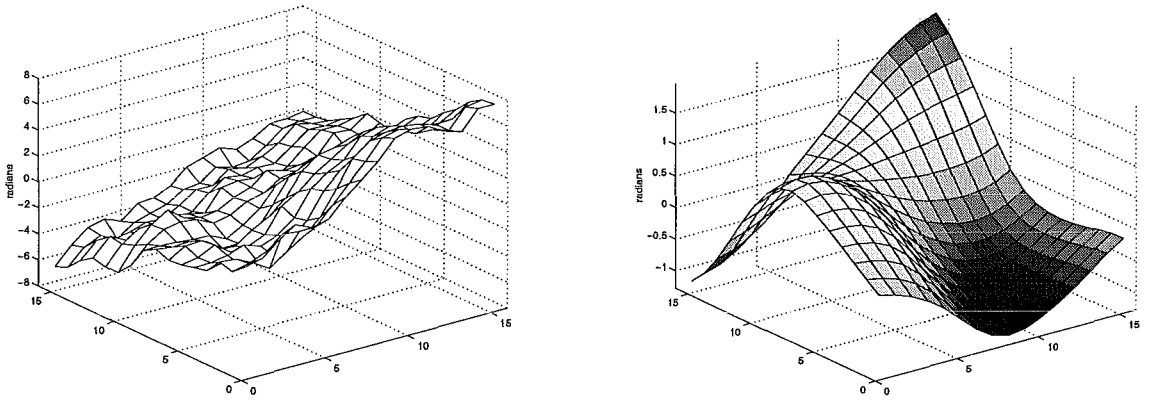
5.6 Multi-aperture Phase Retrieval

In this section two-dimensional phase retrieval with prior information described in section 5.2 is extended to the concept of how phase retrieval can be employed as a robust technique for estimating the wavefront distortion using a lenslet array. This method using a lenslet array is the main principle of the Shack-Hartmann sensor which has been described in the previous chapter. This method shows that the principle problem of phase retrieval, namely divergence is considerably reduced when the aperture is subdivided.

5.6.1 Problem formulation

As noted earlier, the lack of success in using a prior distribution in phase retrieval can in part be attributed to spurious local maxima in the likelihood function caused by the interaction of phase wrapping and the statistical prior. Since nearly all practical phase retrieval algorithms rely on maximizing a cost function of some sort this leads to false

solutions. A typical false solution is shown in Fig. 5.11 where the figure on the left shows the true phase, and the figure on the right a typical false solution produced by a formal Bayesian estimate.



(a) (b)

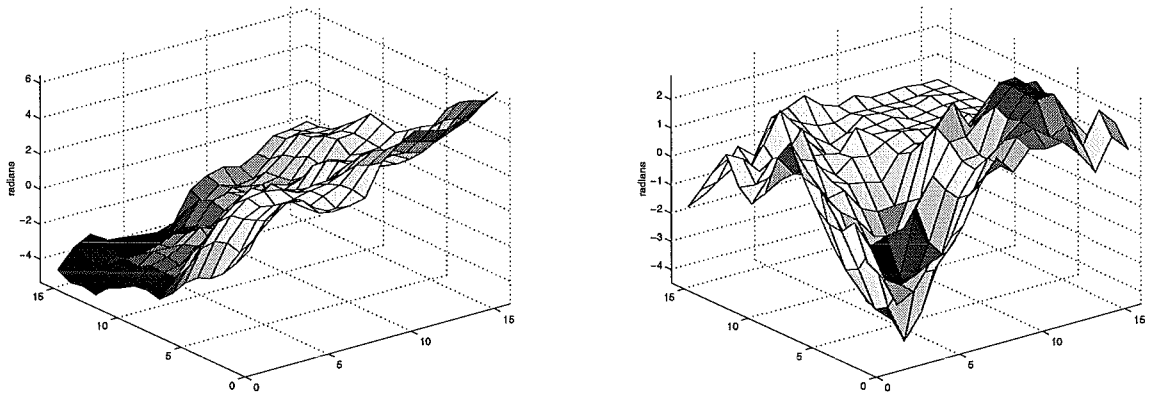
Figure 5.11: *Comparison of the true phase (a) and a false solution (b) caused by interaction of the prior distribution and phase wrapping*

It has been shown in Fig. 5.8c that, apart from the rotational ambiguity, for low levels of phase distortion convergence to a high Strehl solution occurs consistently when the MAP algorithm is used. An obvious solution to the problem of phase retrieval would be to subdivide the aperture in the manner of traditional wavefront sensors. This would also appear to remove the problem of the $\phi(u,v)/-\phi(-u,-v)$ ambiguity. In fact the ambiguity simply recurs in each sub-aperture. This is shown in Fig. 5.12a where the aperture has been subdivided into four smaller apertures. Of the four sub-apertures, three have converged to close to the true solution and one to the rotated solution. This is evident in Fig. 5.12b in which three quadrants of the phase screen show only a small difference between the true and estimated phases. The fourth shows a function equal to twice the even component of the phase in the sub-aperture, i.e.,

$$\phi_s(u,v) + \phi_s(-u,-v) \tag{5.33}$$

where the subscript s indicates the phase in the sub-aperture number s.

A potential problem with the subdivision of the aperture before performing phase retrieval is that subdivision of the aperture effectively creates a number of separate phase problems. The only link between these problems is provided by the prior statistical distribution of the



(a) (b)

Figure 5.12: *Phase retrieval on a subdivided aperture. The reconstruction of the phase in Fig. 5.11a is shown in (a) and the difference of the true phase and the estimated phase in (b). Convergence to the correct phase has occurred in 3 of the 4 sub-apertures with convergence to $-\phi(-u,-v)$ in the final aperture.*

phase, and therefore use of a statistical prior is an essential part of this technique. How a statistical prior is incorporated in the likelihood function is discussed in detail in the section 5.2.

The solution shown in Fig. 5.12b does not correspond to the global maximum of the likelihood function, but it does correspond to a local maximum. The approach taken to overcoming this problem was to start with the initial solution and then cycle through each sub-aperture by replacing the estimated phase with its rotational ambiguity,

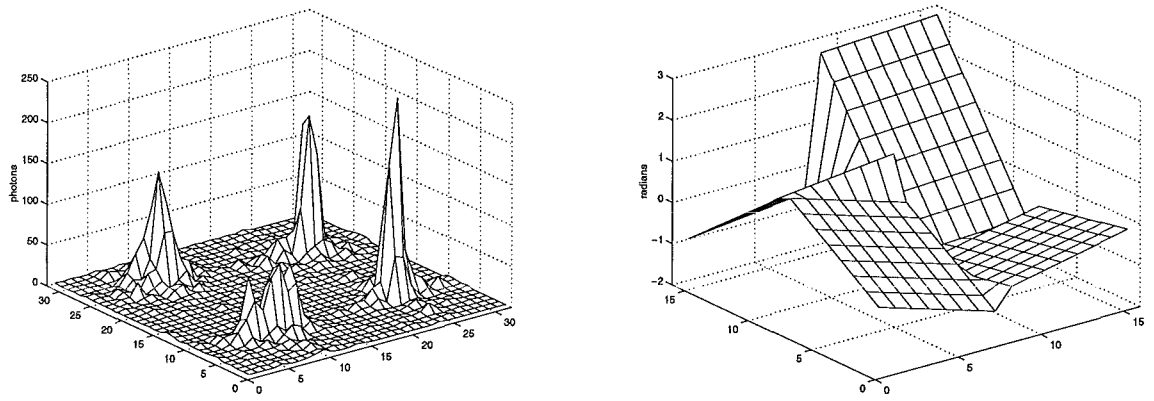
$$\phi_s^{new}(u,v) = -\phi_s(-u,-v) + 2 \times \text{mean } \{\phi_s(u,v)\} \tag{5.34}$$

where a subscript s indicates the current sub-aperture. This was then used as a new starting point for the phase retrieval estimation. If this resulted in a higher likelihood for the resulting phase estimate then this new phase was used as the new starting phase estimate, otherwise the unrotated phase was retained.

Once a cycle of all sub-apertures had been completed, if there was no increase in likelihood observed for any of the new starting estimates the initial phase estimate was taken as the final value, otherwise the whole cycle was restarted.

5.6.2 Results of phase retrieval

Table 5.4 presents the performance of phase retrieval in terms of Strehl ratio for wavefront distortion across a square aperture. The turbulence was generated using Kolmogorov statistics with the ratio of the instrument aperture to r_0 being equal to 4. In all cases where a subdivided aperture was used for phase retrieval, a linear phase starting estimate was formed in each sub-aperture estimated from the centroid of the respective Shack-Hartmann spot (chapter 4). As an illustration, consider a 2×2 Shack-Hartmann speckle image corrupted by image-dependent photon noise shown in Fig. 5.13a. The turbulence is assumed to be $D/r_0 = 4$. The corresponding linear phase starting estimate is depicted in Fig. 5.13b. These linear starting estimates in four sub-apertures were then combined with a piston term for each sub-aperture φ_s, such that the overall starting phase ϕ_j formed by joining the sub-aperture phases minimized $\frac{1}{2}\hat{\mathbf{\Phi}}^T\mathbf{C}^{-1}\hat{\mathbf{\Phi}}$ (Eq. (5.22)).



(a) (b)

Figure 5.13: *(a) A 2×2 simulated noise-corrupted speckle image with $D/r_0 = 4$ and 5000 photons. (b) The corresponding linear phase starting estimate.*

Table 5.4 begins with a raw data which is simulated using the Kolmogorov turbulence with $D/r_0 = 4$. This data produces a mean Strehl of 0.04. When only the slope of the wavefront is estimated, it improves the mean Strehl about 27% resulting a mean Strehl of 0.15. This improvement demonstrates that the tip/tilt contribute a large proportion of the error associated with the aberration. With this removed, there is a significant improvement in the reconstruction of the wavefront. The MAP phase retrieval technique presented in table 5.4 improves further the mean Strehl to 0.34. The local maxima, however, are still present in this result, which are associated with the ambiguities (section 5.5). When the

Phase retrieval technique	Mean Strehl	σ
Raw	0.04	0.010
Slope Estimate	0.15	0.015
Phase retrieval (Full aperture) Ambiguity present	0.34	0.030
Phase retrieval (Full aperture) Ambiguity removed	0.60	0.030
Phase retrieval (4 rad diversity)	0.60	0.030
Phase retrieval (2 x 2 sub-apertures)	0.40	0.010
Phase retrieval (2 x 2 sub-apertures) Iterated	0.67	0.010
Phase retrieval (4 x 4 sub-apertures)	0.79	0.010
Phase retrieval (4 x 4 sub-apertures) Iterated	0.86	0.010

Table 5.4: *Results of phase retrieval for 5000 photons, $D/r_0 = 4$. Results are averaged over 1000 different noise-corrupted speckle images.*

ambiguities are eliminated by using a variable weighting on the prior as discussed in section 5.5.2, the mean Strehl increases to 0.60 (cf. table 5.3). This value is also achieved when a phase diversity technique presented in section 5.5.1 is used. The multi-aperture phase retrieval presented in section 5.6.1 shows significant improvement in Strehl only when an iterative improvement approach is used to eliminate the local maxima. This, however, is only at a very high computational cost. The result of 4×4 multi-aperture phase retrieval achieves a mean Strehl of 0.86 when this approach is used.

5.7 Summary

This chapter has reviewed the iterative phase retrieval algorithms due to Fienup in the first section. These algorithms are based on the earlier method of Gerchberg and Saxton. Fienup has shown that the ER algorithm converges in the sense that the error monotonically decreases. However, in practice the ER algorithm converges very slowly and can stagnate. Of these existing approaches, the HIO remains one of the most effective in overcoming the resulting problems, but can on occasion stagnate.

The second section of this chapter introduces a new approach for the solution of the two-dimensional Fourier phase retrieval problem. It has demonstrated how prior information can be used to improve the phase retrieval process. The introduction of prior information improves the quality of the phase reconstruction in three ways. First, by making the estimate less susceptible to noise on the observed data $d(x, y)$. Second, it improves the correspondence between the functional being maximized and the Strehl ratio. Third, the resultant phase estimate is continuous and does not require phase unwrapping to generate a continuous approximation suitable for phase compensation. It is demonstrated that a full MAP approach is generally superior to a simple smoothness assumption, although the latter does have a computational advantage.

Phase retrieval, even using a MAP approach, remains plagued by local maxima in the likelihood function. However, simulations show that for the important case of phase distortion from Kolmogorov turbulence there exists a region of convergence around the true solution. It is shown that provided the starting estimate of the phase is sufficiently accurate, convergence occurs. Poor starting points can, however, result in convergence to a local maximum.

Two solutions to the problem of local maxima were investigated. The first, a method based on altering the weight on the prior in part overcomes the problems of local maxima. The method relies on forming a progressively more detailed estimate of the phase. While for distortion the convergence to the true solution is still not certain, it is much improved over simple ML or MAP maximization. More importantly, it is shown that using this technique Strehls higher than 0.95 can be achieved.

The second solution relies on subdividing the aperture into smaller sub-apertures in the manner of a Shack-Hartmann sensor. Although subdivision of the aperture effectively creates a number of separate phase problems, an improvement in Strehl can be achieved by an iterative approach but only at a very high computational cost. Moreover, the trapping problem is less relevant than that in full apertures. The numerical analysis shows the effectiveness of the algorithm, even if the presence of noise is quite strong.

The principle drawback of phase retrieval is the relatively long computing time required to find the solution. However, it would not be impossible to obtain real time corrections in the future as computing power increases.

Chapter 6

COMPARATIVE EVALUATION

Wavefront estimation is a critical step in an adaptive optics system. Two main techniques for wavefront estimation are discussed in chapters 4 and 5, respectively. These two algorithms can be formulated using the Bayes theorem but with different models of the relationship between the wavefront and the measurements. The comparative success of these methods for estimating a wavefront phase, measured in the quality of the reconstruction and the computational cost to obtain the reconstruction, is important when one is looking for some methods to solve the wavefront estimation problem in an adaptive optics system.

In this chapter the performance of the Shack-Hartmann sensor with its new analysis presented in chapter 4 is compared with that of the phase retrieval using prior information presented in chapter 5. For convenience, in this chapter the former is simply referred to as Shack-Hartmann sensing, and the latter as phase retrieval. The performance calculations are restricted to common configurations of these two techniques and the fundamental limits imposed by atmospheric effects.

Shack-Hartmann sensing traditionally provided wavefront estimates of a superior quality to those provided by the phase retrieval techniques, and as a consequence is used in practice. One reason for this is that Shack-Hartmann sensing utilizes a priori knowledge which is derived from Kolmogorov turbulence. Section 4.2.3 shows how this information can be expressed in the covariance of the coefficients of a Zernike polynomial expansion. Chapter 5, therefore, investigates the use of the same prior knowledge in phase retrieval, and has

demonstrated a significant improvement.

Phase retrieval is a non-linear problem which is computationally expensive since it requires iterative solution of the equations. Shack-Hartmann sensing, by contrast, simplifies the problem into solving a set of linear equations. Another drawback of the phase retrieval algorithms is related to the existence of false solutions by which the retrieval procedure is trapped. These false solutions do not exist in the Shack-Hartmann sensing. Nevertheless, phase retrieval has the potential to produce a significantly improved performance because of the loss information associated with the linearization of the Shack-Hartmann sensing. In addition, the most significant potential advantage of phase retrieval is related to the fact that it allows larger aperture in the processing of received wavefront, with the corresponding increase of the maximum intensity. However, unless phase retrieval is able to provide a phase estimate of a quality superior to those provided by the Shack-Hartmann sensing, the enormous computational cost of phase retrieval would be unjustified.

In this chapter the Shack-Hartmann sensing is compared with phase retrieval subject to the following conditions. The Strehl ratio which also represents the mean square error (Eq. (3.18)) is used in all comparisons for the following reasons. First, it is mathematically related to the residual error computed from the Shack-Hartmann reconstructions (Eq. (3.50)). Second, it is the simplest method for removing both the effects of the 2π ambiguity and scaling ambiguity in the phase reconstructions, as described in section 5.2. The phase distortion due to the atmospheric turbulence is assumed to be $D/r_0 = 4$ for both algorithms since this has the highest resolution defined by Fried [41]. Also used in this comparison is the number of photons of 5000. The results are summarized in table 6.1 in terms of Strehl ratio. It is, however, important to note that for phase retrieval the achieved Strehls are obtained from the highest possible Strehl it can produce, and thus not from the mean Strehl as presented in table 5.4.

It is obvious in table 5.4 that the results of phase retrieval when local maxima are present are much lower (0.34) than those of the Shack-Hartmann sensing (0.40 or 0.55). However, when the algorithm of increasing the weight of the prior information to remove the local maxima associated with the wrapping ambiguity (section 5.5.2) is used for phase retrieval, a Strehl greater than 0.95 can be achieved as shown in Fig. 5.10. This phase retrieval result is clearly much higher than those of Shack-Hartmann sensing, regardless to the number of sub-apertures. This is simply because phase retrieval utilizes all the available information

Wavefront estimation technique	Maximum Strehl
Raw data	0.04
Phase retrieval Local maxima present	0.34
Phase retrieval Local maxima removed	> 0.95
SHWS (2×2)	0.40
SHWS (4×4)	0.55

Table 6.1: *Results of wavefront estimations in optical telescope for 5000 photons.* $D/r_0 = 4$.

from measurements, whereas Shack-Hartmann sensing only relies on the centroids of the measurements, provided a prior information is used in both cases.

In addition, it is worth noting that the results presented above differ in one important respect. In the Shack-Hartmann sensing Poisson-distributed photon noise is assumed to be generated only by the image spot, and the detector is assumed to have noiseless pixels of infinitely small size. The phase retrieval algorithm, by contrast, has not restricted from the detector noise and the pixel size. Despite these restrictions, the performance of the Shack-Hartmann sensing still cannot exceed that of phase retrieval.

A significant advantage that the Shack-Hartmann sensing has over phase retrieval is relatively small computational cost. To reconstruct a wavefront phase using the Shack-Hartmann sensor was equivalent to $4 \times 200 = 800$ phase retrieval iterations to obtain the result in table 6.1 when local maxima are removed. This is because the weighting of the prior was stepped down in 4 stages, with each stage requiring 200 iterations (section 5.5.2). Another advantage of the Shack-Hartmann sensing is there are no ambiguities in the reconstructions. These ambiguities that result in local maxima do not exist in the Shack-Hartmann sensing simply because the algorithm does not require iterative solution of non-linear equations.

The above results demonstrate that phase retrieval can provide superior improved phase estimates above the Shack-Hartmann sensing, in the situation where a priori information is used in both cases. This superior performance, however, comes at an enormous computational cost and it is questionable whether for high light levels this superiority could be used

to justify the cost of phase retrieval. With the increasing availability of computing power it is possible that this may be the case in the future. At lower light levels, however, the prior is very highly weighted in both algorithms causes the reconstructions to tend to zero. In this case (e.g. 50 photons), as can be seen in table 5.3 and Fig. 4.30, both algorithms achieve almost similar Strehl, i.e. ≈ 0.2.

Chapter 7

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

The purpose of this chapter is to draw together the conclusions that have been drawn from the results presented in chapters 4, 5, and 6. A number of suggestions about how the algorithms could be improved are then made and promising directions for future research in the fields of wavefront estimation are indicated. The conclusions are presented in section 7.1, whilst the suggestions for further research are discussed in section 7.2.

7.1 Conclusions

The problem that originates from the transmission of light through the turbulence atmosphere is discussed in this thesis. Although the atmospheric turbulence affects both the amplitude and the phase, its main effect on the quality of the image is the random phase aberration. Many efforts have been made to compensate the phase distortion so that the quality of the image obtained with ground based telescopes can be improved. Two approaches to estimate the degradation of the wavefronts phase have been discussed, and the conclusions are summarized in this section.

7.1.1 Wavefront sensing

The most common type wavefront sensor in use is the Shack-Hartmann sensor. This sensor linearizes the problem of phase retrieval to solving a linear system of equations, and has been fully described in chapter 4. This section summarizes the main points drawn from that chapter.

A new analysis has been presented which differs from previously published work since the short-exposure spot is used in the calculation of the noise variance rather than the Gaussian approximation. The Gaussian profile has been used extensively in the literature because it is mathematically tractable as well as it matches at the e^{-1} points to the diffraction limited spot. Unfortunately, the resulting noise variance using the Gaussian approximation differs significantly from that of the actual spot. As addressed in section 4.4.1, the derived variance due to photon noise diverges as the size of the detector increases, whilst the conventional analysis shows that the noise variance is not a function of the detector size.

The performance of the sensor in terms of the residual error is therefore shown to be sensitive to the size of the detector surface if the correct function is used. This detector size cannot be too small as this introduces another error due to the truncation of the moving spots. In contrast, if the size of the detector is very large, the photon noise becomes dominant and thus reduces the accuracy of the centroiding. The optimum size has been investigated in terms of these two errors. This size also depends on the number of photons available. If the number of photons is increased the optimum size also increases. This optimum size is of importance since the measurement noise is minimized at this point as shown in section 4.4.1. It is shown that the optimum radius of the detection aperture for $d/r_0 = 1$ is $3.2\beta_{dl}$ for 10 photons, and $6.5\beta_{dl}$ for 100 photons, where β_{dl} is the rms width of the diffraction limited spot.

One of the main factors that motivated the development of this new analysis was the expectation that the performance of the sensor would be less than those expected by previous theoretical work. The simulation results presented in section 4.5.3 demonstrate that the quantitative performance in terms of Strehl ratio shows performance which is lower than those obtained from the conventional analysis. This partly explains the disagreement between theory and actual performance of the sensor.

7.1.2 Phase retrieval in astronomy

A new algorithm that can successfully solve the two-dimensional phase retrieval problem has been described in detail in chapter 5. Conclusions drawn from the theoretical background and simulated results of this chapter are now summarized.

It has been shown that a formal statistical approach can be used to overcome many of the problems which have been encountered in the existing phase retrieval methods. The process of this technique of phase retrieval has been divided into two distinct steps. First, the expression for the covariance of the phase distortion by use of a Kolmogorov model for the turbulence is described. Second, this covariance is employed as part of a formal Bayesian estimate of the phase distortion.

The results presented in section 5.4 demonstrate that the new phase retrieval technique can successfully solve the phase retrieval problem to a certain extent. The algorithm uses an MAP approach to incorporate the statistical properties of the turbulence. This produces consistent solutions to the phase retrieval problem in the absence of contamination. Unfortunately, the algorithm that relies on the gradient search technique, still remains plagued by local maxima in the likelihood function. Nevertheless, there exists a region of convergence around the true solution.

On evaluating the causes of local maxima, it was found that rotational ambiguity of the solution and 2π wrapping were two factors that led into stagnation. A solution to the former is by use of phase diversity. The numerical results, however, show that subdividing the aperture in the manner of traditional wavefront sensor produces a higher Strehl than that of the phase diversity technique. The problem associated with the 2π wrapping is much more complicated. A new method based on increasing the weight on the prior for the initial iterations is proposed, which can be viewed as a form of multi-resolution phase recovery. The result presented in section 5.5 indicate that these two techniques of solving the ambiguities can significantly improve the convergence of iterative phase retrieval algorithms for simulated data.

Finally, in chapter 6 a comparison is made between the two wavefront estimation techniques: the Shack-Hartmann sensing with its new analysis and phase retrieval with prior information. The comparison is based on the quality of the reconstruction in terms of the maximum Strehl ratio achieved and the computational cost. It was found that phase retrieval was

superior to the Shack-Hartmann sensor, and could break the iterative algorithm out of stagnation when the algorithm of increasing the weight of the prior is used. This technique is described in detail in section 5.5.2. The ultimate performance of phase retrieval is able to produce Strehls greater than 0.95 when this technique is used as opposed to $S = 0.55$ obtained from the Shack-Hartmann sensing provided the correct analysis is considered. The result obtained by the Shack-Hartmann sensing is also subject to the assumptions that the detector is noiseless and has an infinite number of pixels.

The successful application of phase retrieval for wavefront estimation encourages its incorporation into adaptive optics systems. Furthermore, the most significant potential advantage of phase retrieval techniques is related to the fact that they allow larger aperture in the processing of the received wavefront, with a corresponding increase of intensity. For these reasons, phase retrieval should become increasingly relevant due to the trend towards higher image reconstruction quality. The only drawback is the high computational cost. In this case, a special purpose hardware would be required. As the computing power increases, the use of phase retrieval in an adaptive optics system could be a practical option at wavelength for which measurement data is unreliable.

7.2 Suggestions for Further Research

Some avenues for future research in the fields of wavefront sensing and phase retrieval are now suggested.

Although the Shack-Hartmann sensor has been a commonly used sensor in the field of astronomical imaging, it can be shown that there still exist several areas in which the performance of the sensor could be improved. One may think of treating the lenslet array dependently from each other instead of separately. This increases the amount of information from each lenslet and hence can improve the estimation of the wavefront.

Furthermore, the wavefront distortion is assumed to be introduced by the Kolmogorov turbulence. In practice, there may be some other contributions to the distortion which are not of Kolmogorov type, such as the turbulence introduced by temperature difference in the dome. It would be feasible to do a complete simulation of an optical telescope incorporating this type of seeing.

An obvious area for future work related to two-dimensional phase retrieval in astronomy is to investigate further whether phase retrieval can be applied as real time technique in an adaptive optics system. This would lead to a new approach to wavefront sensing for adaptive optics systems. It would also be worth looking at the determination of the expansion coefficients of the wavefront in terms of Zernike polynomials. This would thus be to compare with the technique introduced in section 5.2.

Finally, the principle drawbacks of the phase retrieval based wavefront sensing is the relatively long computing time required to find the solution if a general-purpose computer is used. The development of special purpose hardware would be a challenging project.

Bibliography

[1] M. C. Roggemann, B. Welsh, *Imaging through Turbulence*, CRC Press, (1996).

[2] H. W. Babcock, Publ. Astron. Soc. Pac., **65**, (1953).

[3] A. Papoulis, *Probability, random variables, and stochastic processes*, (McGraw-Hill, New York, 1991).

[4] D. M. Alloin and J. -M. Mariotti, *Adaptive Optics for Astronomy*, NATO ASI Series, Kluwer academic publishers, **423**, (1994).

[5] R. G. Lane, "Phase retrieval using conjugate gradient minimization", J. of Modern Opt. **38**, 1797-1813 (1991).

[6] R. Gerchberg, W. Saxton "A practical algorithm for the determination of phase from image and diffraction plane picture", Optik **35**, 237-246 (1972).

[7] J. R. Fienup, C. C. Wackerman, "Phase retrieval stagnation problems and solutions", J. Opt. Soc. Am. A **3**, 1897-1907 (1986).

[8] M. Nieto-Vesperinas, R. Navarro, F. J. Fuentas, "Performance of a simulated-annealing algorithm for phase retrieval", J. Opt. Soc. Am. A **5**, 30-38 (1988).

[9] E. P. Wallner, "Optimal wave-front correction using slope measurements", J. Opt. Soc. Am. A **73**, 1771-1776 (1983).

[10] R. G. Lane, *Blind deconvolution and phase retrieval*, Ph.D thesis, (1988).

[11] T. S. Huang, ed., *Advances in Computer vision and image processing*, **1**, Jai Press Inc., London, (1984).

[12] R. C. Gonzalez, P. Wintz, *Digital Image Processing*, Addison-Wesley, London, (1977).

[13] T. E. Bell, "Electronics and the stars", IEEE Spectrum magazine, August (1995).

[14] J. N. Cederquist, J. R. Fienup, C. C. Wackerman, S. R. Robinson, D. Kryskowski, "Wavefront estimation from Fourier intensity measurements", J. Opt. Soc. Am. A **6**, 1020-1026 (1989).

[15] R. G. Lane, R. Irwan, "Phase retrieval as a means of wavefront sensing", in *Proceedings of International Conference on Image Processing*, B. R. Hunt and R. M. Gray, eds. (IEEE Signal Processing Society Press, Los Alamitos, Calif., 1997), pp. 242-245.

[16] R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics", Opt. Eng. (Bellingham) **21**, 829-832 (1982).

[17] R. G. Paxman, T. J. Schulz, J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity", J. Opt. Soc. Am. A **9**, 1072-1085 (1992).

[18] R. K. Tyson, *Principle of Adaptive Optics*, Academic Press, (1987).

[19] G. A. Tyler, D. L. Fried, "Image-position error associated with a quadrant detector", J. Opt. Soc. Am. A **72**, 804-808 (1982).

[20] C. Witthoft, "Wavefront sensor noise reduction and dynamic range expansion by means of optical image intensification", Opt. Eng. **29**, 1233-1238 (1990).

[21] A. V. Oppenheim, R. W. Schafer, *Discrete-time Signal Processing*, Prentice Hall co., (1989).

[22] A. K. Jain, *Fundamentals of Digital Image Processing*, A Prentice Hall co., (1986).

[23] W. K. Pratt *Digital Image Processing*, Academic Press, Inc., (1991).

[24] H. C. Andrews and B. R. Hunt *Digital Image Restoration*, New Jersey: Prentice Hall, (1977).

[25] J. P. Muller, ed., *Digital Image Processing in Remote Sensing*, Philadelphia: Taylor and Francis, 1998.

[26] V. Krishnamurthi, Y. Liu, T. J. Holmes, "Blind deconvolution of 2D and 3D fluorescent micrographs", Proc. SPIE **1660**, (1992).

[27] B. Friedlander, J. M. Francos, "Model Based Phase Unwrapping of 2-D Signals", IEEE on Signal Processing **44**, (1996).

[28] D. C. Ghiglia, L. A. Romero "Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods", J. Opt. Soc. Am. A **11**, (1996).

[29] M. D. Pritt, J. S. Shipman, "Least-Squares two-dimensional phase unwrapping using FFT's", IEEE Trans. on Image Processing **32**, 706-708 (1994).

[30] M. Born, E. Wolf, *Principles of Optics*, Pergamon Press, 6^{th} edition, (1980).

[31] W. T. Welford, *Optics.* Oxford Science Publications, 3^{rd} edition, (1988).

[32] A. E. Roy, D. Clarke, *Astronomy, principles and practice*, Adam Hilger Ltd., (1982).

[33] J. W. Goodman, *Statistical Optics*, (A Wiley-Interscience publication, 1985).

[34] J. W. Goodman, *Introduction to Fourier Optics*, (McGraw Hill, San Fransisco, Calif., 1968).

[35] B. L. McGlamery, "Computer simulation studies of compensation of turbulence degraded images", Proceedings of SPIE **76**, 225-233 (1976).

[36] R. G. Lane, A. Glindemann, J. C. Dainty, "Simulation of a Kolmogorov phase screen", Waves in Random Media **2**, 209-224 (1992).

[37] C. M. Harding, R. A. Johnston, R. G. Lane, "Fast Simulation of a Kolmogorov phase screen", Appl. Optics, *in press*, (1999).

[38] R. J. Noll, "Zernike polynomials and atmospheric turbulence", J. Opt. Soc. Am. A **66**, 307-310 (1975).

[39] N. Roddier, "Atmospheric wavefront simulation using Zernike polynomials", J. Opt. Soc. Am. A **29**, 1174-1180 (1990).

[40] J. Wang and J. Markay, "Modal compensation of atmospheric turbulence phase distortion", J. Opt. Soc. Am. A **68**, 78-87 (1978).

[41] D. L. Fried, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures", J. Opt. Soc. Am. A **56**, 1372-1379 (1996).

[42] J. Primot, G. Rousset, J. C. Fontanella, "Deconvolution from wavefront sensing: a new technique for compensating turbulence-degraded images", J. Opt. Soc. Am. A **7**, 1598-1608 (1990).

[43] B. M. Welsh, C. S. Gardner, "Performance analysis of adaptive-optics system using laser guide stars and slope sensors", J. Opt. Soc. Am. A **12**, 1913-1923 (1989).

[44] T. Y. Kane, B. M. Welsh, C. S. Gardner, L. A. Thomson, "Wavefront detection optimization for laser guided adaptive telescope", Proc. SPIE, **1114**, 160-171 (1990).

[45] B. M. Welsh, B. L. Ellerbroek, M. C. Roggemann, T. L. Pennington, "Fundamental performance comparison of a Hartmann and a shearing interferometer wavefront sensor", Appl. Opt. **34**, 4186-4195 (1995).

[46] K. A. Winick, "Cramer-Rao lower bounds on the performance of charge-coupled-device optical position estimators", J. Opt. Soc. Am. A **3**, 1809-1815 (1986).

[47] B. L. Ellerbroek, "First-order performance evaluation of adaptive-optics systems for atmospheric-turbulence compensation in extended-field-of-view astronomical telescope", J. Opt. Soc. Am. A **11**, 783-805 (1994).

[48] D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, R. L. White, "Compensation for readout noise in CCD images", J. Opt. Soc. Am. A **12**, 272-283 (1995).

[49] F. Roddier, "Curvature sensing and compensation: a new concept in adaptive optics", Opt. Eng. **29**, 1174-1180 (1998).

[50] M. Tallon, R. Foy, "Adaptive telescope with laser probe: isoplanatism and cone effect", Astron. Astrophys. **235**, 549-557 (1990).

[51] G. A. Tyler, "Rapid evaluation of d_0: the effective diameter of a laser-guide-star adaptive-optics system", J. Opt. Soc. Am. A **11**, 325-338 (1994).

[52] F. Rigaut, E. Gendron, "Laser guide star in adaptive optics: the tilt determination problem", Astron. Astrophys. **261**, 677-684 (1992).

[53] G. Strang, *Linear algebra and its applications*, 3^{rd} edition, (1986).

[54] M. R. Banham, A. K. Katsaggelos, "Digital Image Restoration", IEEE Signal Processing Magazine, 24-41 (1997).

[55] F. Roddier, "The effects of atmospheric turbulence in optical astronomy", in *Progress in Optics*, E. Wolf, ed., (North Holland, Amsterdam, 1981), Vol. 19, pp. 281-376.

[56] D. L. Fried, "Least square fitting of a wavefront distortion estimate to an array of phase difference measurements" J. Opt. Soc. Am. A **67**, 370-375 (1977).

[57] W. H. Southwell, "Wavefront estimation from wavefront slope measurements", J. Opt. Soc. Am. A **70**, 998-1006 (1980).

[58] J. Hermann, "Phase variance and Strehl ratio in adaptive optics", J. Opt. Soc. Am. A **9**, 2257-2258 (1992).

[59] N. F. Law, R. G. Lane, "Wavefront estimation at low light levels", Opt. Commun. **126**, 19-24 (1996).

[60] C. Schwarz, E. Ribak, S. G. Lipson, "Bimorph adaptive mirrors and curvature sensing", J. Opt. Soc. Am. A **11**, 895-907 (1994).

[61] T. J. Schulz, "Multiframe blind deconvolution of astronomical images", J. Opt. Soc. Am. A **10**, 1064-1073 (1993).

[62] J. Phillips, *The NAG library: Programmers Guide*, (Clarendon, Oxford, 1986).

[63] E. Thiebaut, J. -M. Conan, "Strict a priori constraints for maximum likelihood blind deconvolution", J. Opt. Soc. Am. A **12**, 485-492 (1995).

[64] R. G. Lane, R. A. Johnston, R. Irwan, T. J. Connolly, "Regularized blind deconvolution", Technical Digest Series **11**, Optical Society of America, Hawaii, 5-7, (1998).

[65] J. -M. Conan, G. Rousset, P. -Y. Madec, "Wavefront temporal spectra in high-resolution imaging through turbulence", J. Opt. Soc. Am. A **12**, 1559-1570 (1995).

[66] B. R. Hunt, "Matrix formulation of the reconstruction of phase differences from phase differences", J. Opt. Soc. Am. A **69**, 393-399 (1979).

[67] J. R. Fienup, "Phase retrieval algorithms: a comparison", Appl. Opt. **21**, 2758-2769 (1982).

[68] R. P. Millane, "Phase retrieval in crystallography and optics", J. Opt. Soc. Am. A **7**, 394-411 (1990).

[69] R. Irwan and R. G. Lane, "Phase retrieval with prior information", J. Opt. Soc. Am. A **15**, 2302-2311 (1998).

[70] H. Stark, J. Woods, *Probability, Random Processes*, Prentice-Hall, New Jersey, (1994).

[71] A. Levi, H. Stark, "Image restoration by the method of generalized projections with application to restoration from magnitude", J. Opt. Soc. Am. A **1**, 932-943 (1984).

[72] R. H. T. Bates, M. J. McDonnell, *Image restoration and reconstruction*, Clarendon Press, Oxford, (1989).

[73] NASA website, *www.nasa.gov*, (1997).

[74] R. C. Cannon, "Global wave-front reconstruction using Shack-Hartmann sensors", J. Opt. Soc. Am. A **12**, 2031-2039 (1995).

[75] P. A. Bakut, V. E. Kirakosyants, V. A. Loginov, C. J. Solomon, J. C. Dainty, "Optimal wavefront reconstruction from a Shack-Hartmann sensor by use of a Bayesian algorithm", Opt. Commun. **109**, 10-15 (1994).

[76] F. A. Jenkins, H. E. White, *Fundamentals of Optics*, McGraw-Hill, (1957).

[77] R. N. Bracewell, *Two-Dimensional Imaging*, Prentice Hall, (1995).